

00/60/90
jc848 U.S. PRO

06-12-00

A
Bot SLQ

jc836 U.S. PRO
09/591466
00/60/90

"Express Mail" Mailing Label No. EL544961181US Date of Deposit June 9, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, with sufficient postage affixed, and is addressed to **BOX PATENT APPLICATION**, Assistant Commissioner for Patents, Washington, D.C. 20231.

JASON SHINAZI

(Typed or printed name of person mailing paper or fee)

(Signature of person mailing paper or fee)

Patent
Attorney's Docket No. 032266-003

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

**UTILITY PATENT
APPLICATION TRANSMITTAL LETTER**

Box PATENT APPLICATION
Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Enclosed for filing is the utility patent application of Antje von Schaewen for Plant GntI sequences and the use thereof for the production of plants having reduced or lacking N-acetyl glucosaminyl transferase I (GnTI) activity.

Also enclosed are:

- ☒ 6 sheet(s) of ☒ formal ☐ informal drawing(s);
- ☒ a claim for foreign priority under 35 U.S.C. §§ 119 and/or 365 is made to 197 54 622.6 filed in Germany on December 9, 1997;
☒ in the declaration;
- ☐ a certified copy of the priority document;
- ☐ a General Authorization for Petitions for Extensions of Time and Payment of Fees;
- ☐ _____ statement(s) claiming small entity status;
- ☐ an Assignment document;
- ☐ an Information Disclosure Statement; and



21839

[X] Other: Preliminary Amendment ;

[X] Other: Copy of PCT/EP98/08001 and associated papers ;

[X] Other: Copy of PCT/EP98/08001 Translation ;

[X] Other: Sequence Listing ;

[X] An ☐ executed ☒ unexecuted declaration of the inventor(s)
[X] also is enclosed ☐ will follow.

☐ Please amend the specification by inserting before the first line the sentence --This application claims priority under 35 U.S.C. §§119 and/or 365 to filed in on ; the entire content of which is hereby incorporated by reference.--

☐ A bibliographic data entry sheet is enclosed.

[X] The filing fee has been calculated as follows [X] and in accordance with the enclosed preliminary amendment:

C L A I M S					
	NO. OF CLAIMS		EXTRA CLAIMS	RATE	FEE
Basic Application Fee					\$690.00 (101)
Total Claims	20	MINUS 20 =	0	x \$18.00 (103)	0
Independent Claims	3	MINUS 3 =	0	x \$78.00 (102)	0
If multiple dependent claims are presented, add \$260.00 (104)					0
Total Application Fee					\$690.00
If verified Statement claiming small entity status is enclosed, subtract 50% of Total Application Fee					
Add Assignment Recording Fee \$40.00 (581) if Assignment document is enclosed					
TOTAL APPLICATION FEE DUE					\$690.00

☐ This application is being filed without a filing fee. Issuance of a Notice to File Missing Parts of Application is respectfully requested.

- ☒ A check in the amount of \$ 690.00 is enclosed for the fee due.
- ☐ Charge \$ _____ to Deposit Account No. 02-4800 for the fee due.
- ☒ The Commissioner is hereby authorized to charge any appropriate fees under 37 C.F.R. §§ 1.16, 1.17 and 1.21 that may be required by this paper, and to credit any overpayment, to Deposit Account No. 02-4800. This paper is submitted in duplicate.

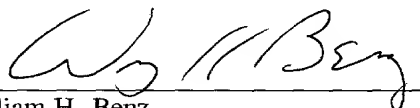
Please address all correspondence concerning the present application to:

William H. Benz
Burns, Doane, Swecker & Mathis, L.L.P.
P.O. Box 1404
Alexandria, Virginia 22313-1404.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

Date: June 9, 2000

By: 
William H. Benz
Registration No. 25,952

P.O. Box 1404
Alexandria, Virginia 22313-1404
(650) 622-2300

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of)
)
von Schaewen, Antje) Group Art Unit: Unassigned
)
Application No.: Unassigned) Examiner: Unassigned
)
Filed: June 9, 2000)
)
For: Plant GntI sequences and the use)
 thereof for the production of plants)
 having reduced or lacking N-acetyl)
 glucosaminyl transferase I (GnTI))
 activity)

PRELIMINARY AMENDMENT
PURSUANT TO MPEP 714.09

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Prior to calculating the filing fee in this application, please amend this application as follows:

In the Specification

Page 1, before line 5 add the following:

— This is a continuation of Patent Cooperation Treaty application EP98/08001. That PCT application was filed on December 9, 1998 and designated the United States of America and additional countries. That PCT application is hereby incorporated by reference in its entirety. —

In the Claims

Please cancel claims 1 and 4-30.

Please add claim 31

- 31. A method for the production of glycoproteins displaying minimal, uniform GlcNac₂Man₃-residues, comprising cultivating a transgenic plant, parts of

transgenic plants or transformed plant cells, and isolating the desired glycoprotein from the material cultivated, characterized in that the transgenic plant, parts of transgenic plants or transformed plant cells, respectively, is/are transformed with an antisense construct or a sense construct, comprising an antisense DNA or a sense DNA with respect to the DNA sequence for a gene or a cDNA for plant N-acetyl glucosaminyl transferase I or a part thereof, for elimination or reduction of the activity of said N-acetyl glucosaminyl transferase, wherein the antisense or sense construct optionally contains additional regulatory sequences for the transcription of the respective antisense or sense DNA. —

In claim 2, line 1, change "claim 1" to — claim 31—.

Please add claims 32-48.

- 32. The method according to claim 31, characterized in that the transgenic plant used is additionally transformed with the gene encoding the desired glycoprotein.
33. The method according to claim 2, characterized in that the transgenic plant used is additionally transformed with the gene encoding the desired glycoprotein.
34. The method according to claim 3, characterized in that the transgenic plant used is additionally transformed with the gene encoding the desired glycoprotein.
35. An isolated DNA, comprising a DNA molecule encoding a sequence or the complementary thereof, which is selected from the group consisting of:
- SEQ ID NOs:1, 3 and 5;
 - a DNA sequence encoding the amino acid sequence of SEQ ID Nos: 2, 4 or 6;
 - a DNA sequence which hybridizes under stringent conditions to SEQ ID NOs:1, 3 or 5, or the complementary thereof; and

a DNA sequence which hybridizes under stringent conditions to a DNA sequence, or the complementary thereof, which encodes SEQ ID NOs:2, 4 or 6.

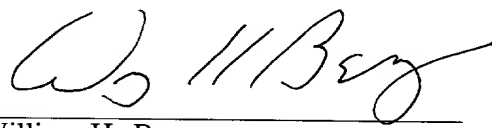
36. An isolated DNA which encodes a polypeptide having N-acetyl glucosaminyl transferase I activity and which hybridizes under stringent conditions to a DNA of claim 35.
37. A DNA construct comprising the DNA of claim 35 in the sense or anti-sense orientation.
38. A DNA construct comprising the DNA of claim 36 in the sense or anti-sense orientation.
39. A microorganism transformed with the DNA construct of claim 37.
40. A microorganism transformed with the DNA construct of claim 38.
41. A protein encoded by the DNA of claim 35.
42. A protein encoded by the DNA of claim 36.
43. An antigen, characterized in that it comprises:
 - the amino acid sequence given in SEQ ID NO: 2, SEQ ID NO: 4 or SEQ ID NO: 6, or
 - amino acids 74 to 446 of the amino acid sequence given in Fig. 2, or
 - an amino acid sequence derived from the amino acid sequences given in SEQ ID NO: 2, 4 or 6 by substitution, deletion, insertion and/or modification of individual amino acids and/or smaller groups of amino acids,with the proviso, that upon immunization of a host with the antigen, said antigen raises an immunological reaction, including the production of antibodies directed against the antigen.
44. A monoclonal or polyclonal antibody, characterized in that it specifically recognizes and binds at least one protein of claim 41.
45. A monoclonal or polyclonal antibody, characterized in that it specifically recognizes and binds at least one protein of claim 42.

46. A monoclonal or polyclonal antibody, characterized in that it specifically recognizes and binds at least one antigen of claim 43.
47. A transgenic plant, transgenic seed, transgenic reproduction material, part of a transgenic plant or transformed plant cell, obtainable by integration of one or more antisense or sense DNA of claim 35 under the control of a promoter effective in plants, into the genome of a plant, or by viral infection by means of a virus containing one or more antisense or sense DNA of claim 35, for an extrachromosomal propagation and transcription of the antisense construct(s) in the plant tissue infected.
48. A transgenic plant, transgenic seed, transgenic reproduction material, part of a transgenic plant or transformed plant cell, obtainable by integration of one or more antisense or sense DNA of claim 36 under the control of a promoter effective in plants, into the genome of a plant, or by viral infection by means of a virus containing one or more antisense or sense DNA of claim 36, for an extrachromosomal propagation and transcription of the antisense construct(s) in the plant tissue infected. —

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

By:



William H. Benz
Registration No. 25,952

P.O. Box 1404
Alexandria, Virginia 22313-1404
(650) 622-2300

Date: June 9, 2000

Date of Deposit June 9, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, with sufficient postage affixed, and is addressed to BOX PATENT APPLICATION, Assistant Commissioner for Patents, Washington, D.C. 20231.

JASON SHINAZI

(Typed or printed name of person mailing paper or fee)

(Signature of person mailing paper or fee)

Patent

Attorney Docket No. 032266-003

Plant GntI sequences and the use thereof for the production
of plants having reduced or lacking N-acetyl glucosaminyl
transferase I (GnTI) activity

5 The present invention relates to plant GnTI sequences, in particular, plant nucleic acid sequences encoding the enzyme N-acetyl glucosaminyl transferase I (GnTI), as well as GntI
10 antisense or sense constructs, deduced therefrom, and their translation products, antibodies directed against said translation products as well as the use of the sequence information for the production of transformed microorganisms and of
15 transgenic plants, including those with reduced or lacking N-acetyl glucosaminyl transferase I activity. Such plants with reduced or lacking N-acetyl glucosaminyl transferase I activity are of great importance for the production of glyco-
20 proteins of specific constitution with respect to their sugar residues.

Prior art:

25 In eukaryotes, glycoproteins are cotranslationally assembled in the endoplasmatic reticulum (ER) (i.e. during import into the ER lumen) by the attachment of initially membrane bound
30 glycans (via dolichol pyrophosphate) to specific asparagine residues in the growing polypeptide chain (N-glycosylation). In higher organisms, sugar units located at the surface of the folded polypeptide chain are subjected to further trimming and
35 modification reactions (ref. 1) in the Golgi cisternae. Initially, typical basic Glc₃Man₉GlcNAc₂ units of the high-mannose type are formed by means of different glycosidases and glycosyl transferases in the ER, which during the passage
40 through the different Golgi cisternae are subsequently converted to so-called complex glycans. The latter are characterized by less mannose units and the presence of

additional sugar residues, such as fucose, galactose and/or xylose in plants and sialic acid (N-acetyl neuraminic acid, NeuNAc) in mammals (ref. 1,2,3). The extent of the modifications can differ between glycoproteins. Single polypeptide chains may carry heterogeneous sugar chains. Furthermore, the glycosylation pattern may vary for a specific polypeptide (tissue specific differences), and does not always have to be uniform with respect to a specific glycosylation site, which is referred to as microheterogeneity (ref. 4,5). Up to now, the role of asparagine bound glycans is barely understood, which i.a. results from the fact, that said glycans may serve several functions (ref. 6). However, it can be assumed, that e.g. protection of a polypeptide chain from proteolytic degradation can also be achieved by glycans of a more simple oligomannosyl type (ref.7).

Description of problems:

Glycoproteins are highly important in medicine and research. However, large scale isolation of glycoproteins is time-consuming and expensive. The direct use of glycoproteins isolated conventionally often raises problems, since upon administration as a therapeutic, single residues of the glycan components may cause undesired side effects. In this context, the glycan component predominantly contributes to the physico-chemical properties (such as folding, stability and solubility) of the glycoproteins. Furthermore, isolated glycoproteins, as already mentioned above, rarely carry uniform sugar residues, which is referred to as microheterogeneity.

For the production of glycoproteins for medicine and research, yeasts prove to be unsuitable, since they are only able to perform glycosylations of the so-called high-mannose type. While insects and higher plants exhibit complex glycoprotein modifications, these, however, differ from those of animals. Therefore, glycoproteins isolated from plants have a strong antigenic effect in mammals. In most cases, animal organisms

with glycosylation defects are not viable, since terminal glycan residues (e.g. of membraneous glycoproteins) mostly possess biological signal function and are indispensable for cell-cell-recognition during the course of embryogenesis. Mammalian cell lines with defined glycosylation defects already exist, the cultivation of which, however, is labour-intensive and expensive.

For mammals, different glycosylation mutants have been described in detail at the cell culture level (ref. 7,8,9,10). Said mutants are either defective in biosynthesis of mature oligosaccharide chains attached to dolichol pyrophosphate or in glycan processing or show alterations in their terminal sugar residues, respectively. Some of these cell lines exhibit a conditional-lethal phenotype or are defective in intracellular protein transport. The consequences of said defects for the intact organism are difficult to estimate. It has been observed, that a modification in the pattern of complex glycans on the cell surfaces of mammals is accompanied by the formation of tumours and metastases, although a functional relationship could not yet unambiguously be demonstrated (ref. 9). Therefore, in mammals glycosylation mutants are very rare. These defects, summarized under HEMPAS (Hereditary Erythroblastic Multinuclearity with a Positive Acidified Serum lysis test) (ref. 10,11), are based either on a deficiency in mannosidase II and/or low levels of the enzyme N-acetyl glucosaminyl transferase II (GnTII), and have strongly limiting effects on the viability of the mutated organism. *GntI* knock-out mice, in which the gene for GnTI has been destroyed, already die *in utero* of multiple developmental defects (personal communication, H. Schachter, Toronto).

Until recently, no comparable mutants were known for plants. By the use of an antiserum, which specifically recognizes complex modified glycan chains of plant glycoproteins and which predominantly is directed against the highly antigenic

5 β 1 \rightarrow 2 linked xylose residues (ref. 12), the applicant was able to isolate several independent mutants from an EMS mutagenized F2 population of the genetic model plant *Arabidopsis thaliana*, which no longer showed complex glycoprotein modification (complex glycan, *cgl* mutants). After at least five backcrosses, each followed by intermittent selfings (to screen for the recessive defects), the glycoproteins were analyzed. These glycoproteins mainly carried glycans of the $\text{Man}_5\text{GlcNAc}_2$ type, indicating a defect in N-acetyl glucosaminyl transferase I (GnTI) (ref. 8). Indeed, the *Arabidopsis cgl* mutants lacked GnTI activity (ref. 13), which normally catalyzes the first reaction in the synthetic pathway to complex modified sugar chains (ref. 1). However, according to observations so far, the viability of the mutated plants is not affected. In recent publications, plants are suggested as a putative source for the production of pharmaceutically relevant glycoproteins or vaccines (ref. 14,15). There however, it was overlooked, that glycoproteins isolated from plants may cause severe immune reactions in mammals, which up to now obstructed the production of heterologous glycoproteins in cultivated plants.

25 The applicant could demonstrate by way of example for the *Arabidopsis cgl* mutant, that plants can manage without complex modified glycoproteins to a great extent (ref. 13). Initially, secretory proteins are normally glycosylated in the ER of the mutant. In the Golgi apparatus of the *cgl* mutant, however, the oligomannosyl chains linked to the polypeptide backbone via asparagine residues (N-glycosylation) then remain at the stage of $\text{Man}_5\text{GlcNAc}_2$ residues, since N-acetyl glucosaminyl transferase I (GnTI) activity is missing (Fig. 1). By this biosynthesis block, the plant specific complex glycoprotein modification and in particular the attachment of α 1 \rightarrow 3 fucose and β 1 \rightarrow 2 xylose residues is prevented, whereby the strong antigenic effect on the mammalian organism is absent. However, *Arabidopsis* as a herb only has little utilizable biomass. Therefore, for the large scale production of biotechnologically relevant glycoproteins these *cgl* plants are less

suitable. As an alternative, cultivars, especially Solanaceae, such as potato, tobacco, tomato or pepper and furthermore alfalfa, canola, beets, soybean, lettuce, corn, rice and grain, with missing or highly reduced GnTI activity, would be ideal for the production of heterologous glycoproteins in plants. For this purpose, the methods of homology-dependent gene silencing would be applicable (ref. 16, 17).

As Fig. 3 demonstrates, the homology of the first determined plant *GntI* sequence from potato (*Solanum tuberosum* L., St) is extraordinary low in comparison to the corresponding known sequences of animal organisms (only 30-40% identity at the protein level, cf. Fig. 3A). Therefore, by the use of heterologous *GntI* gene sequences an efficient reduction of endogenous complex glycoprotein modification in plants by means of antisense or sense suppression, respectively, (ref. 21), probably cannot be achieved.

Thus, in medicine and research there is still the need for a cost-effective production in suitable organisms of recombinant glycoproteins with a minimum of uniform, i.e. defined sugar residues.

Nature of the present invention:

Since the applicant for the first time has been able to isolate and elucidate plant *GntI* cDNA sequences, it is now possible i.a. to obtain and, in particular, to generate any plant having reduced or missing GnTI activity, and to detect the corresponding mutants, respectively, by means of reverse genetic approaches following transposon (ref. 18) or T-DNA insertion (ref. 19), respectively, so as to produce glycoproteins with low antigenic potential in said mutants.

i) Enzymes

Generally, the present invention comprises different N-acetyl glucosaminyl transferase I enzymes (EC 2.4.1.101) from plants, e.g. potato (*Solanum tuberosum* L.), tobacco (*Nicotiana tabacum* L.) and *Arabidopsis thaliana* (L.). In particular, the present invention relates to enzymes, which exhibit or contain the amino acid sequences given in Fig. 2 and 3B as well as in the accompanying sequence protocol.

Further, the invention comprises enzymes, which are derived from amino acid sequences of the above mentioned enzymes by amino acid substitution, deletion, insertion, modification or by C-terminal and/or N-terminal truncation and/or extension, and which - if showing enzymatic activity - exhibit a specificity comparable to that of the starting enzyme, i.e. N-acetyl glucosaminyl transferase I activity, and optionally a comparable activity.

In the present context, by the term "comparable activity" an activity is understood, which is in the range of up to 100% above or below that of the starting enzyme. Accordingly, also comprised by the invention are derived enzymes or proteins with very low or completely lacking enzymatic activity, which is detectable by means of one or more of the tests mentioned as follows. The enzyme activity is determined by a standard assay, which is performed with microsomal fractions either under radioactive conditions, e.g. using UDP-[6-³H]GlcNAc as a substrate (ref. 13) or non-radioactive conditions (HPLC method; ref 20). Plant GnTI activity can be detected on the subcellular level in Golgi fractions (ref. 21). On account of low yields, however, it is almost impossible to enrich the enzyme from plants.

Alternatively, an enzyme derived according to the present invention, may optionally be defined as an enzyme, for which a DNA sequence encoding the enzyme can be determined or

derived, which hybridizes to a DNA sequence encoding the starting enzyme or to a complementary sequence under stringent conditions, as defined as follows.

5 For example, an enzyme derived in such a manner represents an isoform, which comprises the amino acids 74 to 446 of the amino acid sequence illustrated in Fig. 2 and in SEQ ID No:1 and 2. This isoform *i.a.* lacks the membrane anchor formed by amino acids 10 to 29. As a result, this enzyme isoform may be located
10 in the plant cytosol.

As examples for C- and/or N-terminally extended proteins, fusion proteins can be mentioned, comprising in addition to
15 an amino acid sequence according to the invention a further protein, which e.g. exhibits a different enzymatic activity or which may be easily detected in another manner, such as by means of fluorescence or phosphorescence or on account of a reactivity with specific antibodies or by binding to suitable
20 affinity matrices.

Furthermore, the invention comprises fragments of said enzymes, which optionally no longer exhibit any enzymatic activity. Generally, these fragments show an antigenic effect in a host immunized with said fragments, and may accordingly be employed
25 as an antigen for the production of monoclonal or polyclonal antibodies by immunization of a host with those fragments.
30

Moreover, this invention also relates to N-acetyl glucosaminyl transferase I enzymes from other varieties and plant species, which are obtainable on account of hybridization of their genes or one or more regions of their genes:
35

- to one or more of the DNA sequences and/or DNA fragments of the present invention, as discussed below and/or
40
- to suitable hybridization probes according to the invention, which may be prepared on the basis of the amino acid sequences mentioned in the sequence protocol considering the degeneration of the genetic code.

Further comprised by the invention in accordance with the above are enzymes or proteins derived from these N-acetyl glucosaminyl transferase I enzymes, including fusion proteins thereof, as well as fragments of all of these enzymes or proteins.

ii) Antibodies

Another aspect of the present invention relates to the use of the amino acid sequences mentioned above and of fragments thereof having antigenic effects, respectively, for the production of monoclonal or polyclonal antibodies or antisera by immunizing hosts with said amino acid sequences or fragments, respectively, as well as of antibodies or antisera, respectively, *per se*, which specifically recognize and bind to the enzymes and/or antigens described above. The general procedure and the corresponding techniques for the generation of polyclonal and monoclonal antibodies are all well-known to the persons skilled in the art.

Exemplarily, by the use of a fragment of the *GntI* cDNA (nucleotides 275 to 1395) represented in Fig. 2 and SEQ ID NO: 1, the recombinant GntI protein from *Solanum tuberosum* with 10 N-terminal histidine residues (His-tag) was overexpressed in *E. coli*, and, following affinity purification via a metal-chelate matrix, was employed as an antigen for the production of polyclonal antisera in rabbits (cf. Examples 5 and 6).

One possible use of the antibodies of the invention resides in the screening of plants for the presence of N-acetyl glucosaminyl transferase I.

Binding of the antibody according to the present invention to plant protein(s) indicates the presence of N-acetyl glucosaminyl transferase I enzyme detectable with said antibody. In

general, this antibody may then be covalently bound to a carrier in a later step, and optionally be employed for the enrichment or purification of the enzyme by means of column chromatography.

5 On the other hand, a negative binding result using the antibody of the present invention, i.e. lack of binding to the plant proteins, may suggest, that N-acetyl glucosaminyl transferase I enzyme is absent (or highly modified by mutation), and thus, that N-acetyl glucosaminyl transferase I activity of a plant investigated is missing or highly reduced.

10 Techniques for the realization of the screening assays mentioned above or the enrichment or purification of enzymes by the use of antibody columns or other affinity matrices (cf. Examples 5 and 6) are well-known to those skilled in the art.

15
20 iii) DNA sequences

25 The present invention further comprises DNA sequences encoding the amino acid sequences of the invention, including amino acid sequences derived therefrom according to the above provisions. In particular, the invention relates to the respective gene, which is the basis of the amino acid sequences described in the Figures 2 and 3B and the sequence protocol, and especially, to the cDNA sequences described in Fig. 2 and the sequence protocol, as well as to DNA sequences derived from these genes and DNA sequences.

30
35 By the term "derived DNA sequences" are meant sequences, which are obtained by substitution, deletion and/or insertion of one or more and/or smaller groups of nucleotides of the sequences mentioned above and/or by truncation or extension
40 at the 5' and/or 3' terminus. Modifications within the DNA sequence may lead to derived DNA sequences, which encode amino acid sequences being identical to the amino acid sequence encoded by the starting DNA sequence, or to such

sequences, in which, compared to the amino acid sequence, which is encoded by the starting DNA sequence, single or a few amino acids are altered, i.e. substituted, deleted and/or inserted, as well as to such sequences, which - optionally in addition - are truncated and/or extended at the C-terminus and/or N-terminus.

Furthermore, the present invention also extends to the complementary sequences of the genes and DNA sequences according to the invention, as well as the RNA transcription products thereof.

Particularly comprised by the present invention are all sequences derived according to the above provisions, which over their entire length or only with one or more partial regions hybridize under stringent conditions to the starting sequences mentioned above or to the sequences complementary thereto or to parts thereof, as well as DNA sequences comprising such sequences.

By the term "hybridization under stringent conditions" in the sense of the present invention is understood a hybridization procedure according to one or more of the methods described below. Hybridizing: up to 20 h in PEG buffer according to Church and Gilbert (0.25 M Na_2HPO_4 , 1mM EDTA, 1% (w/v) BSA, 7% (w/v) SDS, pH 7.5 with phosphoric acid; ref. 22) at 42°C or in standard hybridization buffers with formamide at 42°C or without formamide at 68°C (ref. 23). Washing: 3 times at 65°C for 30 min in 3x SSC buffer (ref. 23), 0.1% SDS.

In the sense of the present application, the term "hybridization" always means hybridization under stringent conditions, as mentioned above, even if this is not explicitly indicated in the individual case.

Moreover, the invention relates to fragments of the DNA sequences mentioned above, including the DNA sequences

derived in accordance with the above provisions, to fragments derived from such fragments by nucleic acid substitution, insertion and/or deletion as well as the corresponding fragments with sequences complementary thereto. Such fragments are *i.a.* suitable as sequencing or PCR primers, screening probes and/or for uses as discussed below. For the use as a screening or hybridizing probe, the DNA fragments according to the present invention are frequently employed as radio-labelled fragments. Fragments carrying sequences, which are derived from the starting sequences defined above by substitution, deletion and/or insertion of one or more nucleotides, and the sequences complementary thereto, respectively, are comprised by the invention to that extent, as said fragments hybridize under the above mentioned stringent conditions to the starting sequences, or to the sequences complementary thereto, respectively.

On the basis of the DNA sequences mentioned in the sequence protocol and in Figure 2, DNA fragments according to the invention may for example be obtained starting from plant DNA by means of restriction endonucleases using appropriate restriction sites or by employment of PCR by means of primers appropriately synthesized, or may, as an alternative, also be chemically synthesized. Such techniques are well-known to those skilled in the art.

Moreover, the invention relates to any DNA sequences, which represent a gene or are a part of a gene encoding the enzyme N-acetyl glucosaminyl transferase I and, which in their entirety or in a partial region thereof hybridize under stringent conditions

- to one or more of the DNA sequences of the invention and/or
- to one or more of the DNA fragments of the invention and/or
- to a DNA sequence, which is derived from the amino acid sequences mentioned in the sequence protocol considering the degeneration of the genetic code.

For this purpose, hybridization or screening probes are used as DNA fragments, which generally comprise at least 15 nucleotides, typically between 15 and 30 nucleotides, and, if necessary, substantially more nucleotides. As an example, the primers employed in Example 1 may be used. Alternatively, DNA sequences of appropriate length, derived from the DNA sequences mentioned in the sequence protocol, may be used. As a third possibility, appropriate hybridization probes according to the invention may be developed starting from the amino acid sequences mentioned in the sequence protocol considering the degeneration of the genetic code.

In this respect, a subject-matter of the present invention are also genes encoding N-acetyl glucosaminyl transferase I, which may be detected from other varieties or plant species on account of the hybridization thereof to above mentioned hybridization probes, as well as DNA sequences, DNA fragments and constructs, which are derived therefrom in accordance with the above provisions.

The isolation of the corresponding gene and sequencing thereof following detection by means of the hybridization probes of the invention are well within the skills of a specialist in this field, and are detailed by way of example with respect to N-acetyl glucosaminyl transferase I from *Solanum tuberosum* and to the corresponding enzymes from *Nicotiana tabacum* and *Arabidopsis thaliana* (partial sequence) in the examples.

Finally, another subject matter of the present invention are antisense sequences with respect to any of the above DNA sequences.

iv) Constructs

Also comprised by the invention are constructs, which may optionally comprise besides additional 5' and/or 3' sequences, e.g. linkers and/or regulatory DNA sequences or other

modifications, the DNA sequences of the invention, including the DNA sequences derived as detailed above.

5 An example for this are hybridization or screening probes, which in addition to a DNA sequence of the invention also comprise a detection agent for the verification of hybridization products, which in this case typically is non-radioactive, e.g. fluorescent or phosphorescent molecules, 10 biotin, biotin derivatives, digoxigenin and digoxigenin derivatives. In this respect, however, radioactive or non-radioactive detection agents may be considered, which may be attached to the DNA sequence according to the present invention e.g. by means of end labelling. 15

A subject-matter of the invention are also antisense and sense constructs with respect to the DNA sequences and fragments according to the present invention, i.e. with respect to 20

- the DNA sequences mentioned in the sequence protocol and the corresponding genes;
- 25 - the DNA sequences derived therefrom in accordance with the above provisions;
- one or more regions of these DNA sequences;
- DNA sequences, especially from other varieties or plant species, which represent a gene or are a part of a gene, 30 encoding the enzyme N-acetyl glucosaminyl transferase I; and which hybridize under stringent conditions
 - to one or more of the above DNA sequences and/or
 - 35 -- to one or more of the above DNA fragments and/or
 - to a DNA sequence, which is derived from the amino acid sequences mentioned in the sequence protocol considering the degeneration of the 40 genetic code.

Furthermore, the present invention extends to any DNA-transfer systems such as vectors, plasmids, viral and phage genomes or cosmids, which contain the DNA sequences according

to the present invention, e.g. the *GntI* gene, cDNA and DNA regions according to the invention, as mentioned in the sequence protocol, fragments thereof, in particular antisense or sense constructs and/or cDNA sequences derived therefrom according to the above provisions.

Various techniques for the production or synthesis of DNA, DNA fragments, constructs and transfer systems according to the invention, e.g. digestion by means of restriction endonucleases, PCR amplification using suitable primers, optionally followed by cloning and additional chemical or enzymatic modification starting from plant DNA are well-known to those skilled in the art.

One possibility of application of the DNA hybridization probes according to the invention is the detection of N-acetyl glucosaminyl transferase I genes in plants other than those, from which the DNA sequences mentioned in the sequence protocol were obtained, or the detection of potential (other) isoforms of the N-acetyl glucosaminyl transferase I gene in the starting plants *Solanum tuberosum*, *Nicotiana tabacum* and *Arabidopsis thaliana*.

If it is possible to make use of a plant genomic library or cDNA library for the hybridization experiment, a positive hybridization result of such screening of each library may indicate a clone or a few clones, which contain the desired sequence completely or in part, i.e. the N-acetyl glucosaminyl transferase I gene, combined with only a limited amount of other DNA from the genome of the target plant, which appropriately facilitates cloning and sequencing of the target gene. As an alternative, a PCR amplification of the gene or parts thereof may also be carried out starting from plant DNA and suitable constructs, so-called PCR primers, to facilitate cloning and sequencing.

One use of sequencing primers of the invention, which are synthesized starting from suitable regions of the sequences according to the invention, e.g. enables genomic sequencing starting from the entire target plant genomic DNA cleaved by restriction endonucleases, by means of the Church-Gilbert technique, as well as sequencing at the cDNA level following RT-PCR amplification of the total RNA of the target plant (cf. Expl. 1).

An alternative possibility of application of the DNA hybridization probes according to the present invention derived from the DNA sequences mentioned in the sequence protocol, is the use thereof according to the invention for the detection of plants with reduced or lacking N-acetyl glucosaminyl transferase I activity. The hybridization experiment serves to detect the N-acetyl glucosaminyl transferase I (*GntI*) gene by which it may be concluded, e.g. owing to a negative hybridization result under stringent conditions, that the *GntI* gene, and thus, N-acetyl glucosaminyl transferase I activity in a plant investigated is lacking.

Such hybridization techniques for the detection of proteins or genes particularly in plant material by means of DNA probes are also known to the persons skilled in the art. In this context, it is referred to the above statements under item iii) for possible hybridization conditions. Generally, suitable DNA hybridization probes comprise at least 15 nucleotides of a sequence, which for example is derived from the cDNA sequences mentioned in Fig. 2 and the sequence protocol or from the corresponding *GntI* genes.

v) Transformed microorganisms

Furthermore, the invention relates to microorganisms, such as bacteria, bacteriophages, viruses, unicellular eukaryotic organisms, such as fungi, yeasts, protozoa, algae, and human, animal and plant cells, which have been transformed by one or

more of the DNA sequences of the invention or one or more of the constructs of the invention, as illustrated above.

Transformed microorganisms according to the present invention are used e.g. as expression systems for the transforming foreign DNA to obtain the corresponding expression products. For this purpose, typical microorganisms are bacteria, e.g. such as *E. coli*. Furthermore, transformed microorganisms according to the invention, in particular agrobacteria, may be employed e.g. for the transformation of plants by transmission of the transforming foreign DNA.

Methods for the transformation of cells of microorganisms by (foreign) DNA are well-known to those skilled in the art.

For this purpose, e.g. constructs referred to as expression vectors are used, which contain the DNA sequence of the invention under control of a constitutive or inducible promoter, which, if necessary, is additionally tissue specific, so as to enable the expression of the introduced DNA in the target or host cell.

Therefore, a further aspect of the invention is a method for the production of the enzymes and proteins of the invention by using one or more of the transformed microorganisms of the present invention. The method comprises cultivating at least one microorganism transformed by the DNA of the invention, in particular by one of the cDNAs mentioned in the sequence protocol, under the control of an active promoter, as defined above, and isolating the enzyme of the invention from the microorganisms, and, if applicable, also from the culture medium. It is understood, that this method also relates to the production of enzymes and proteins, respectively, which are derived from the enzymes according to the present invention from *Solanum tuberosum*, *Nicotiana tabacum* and *Arabidopsis thaliana*, as defined under i) above.

Methods for the cultivation of transformed microorganisms are well-known to those skilled in the art. For example, the isolation of the expressed enzyme may be employed according to the method described in Example 5 by means of metal-chelate chromatography or, alternatively, by chromatography via columns, which contain the antibodies against the enzyme bound to the packing material.

vi) Transgenic plants

Furthermore, the invention comprises transgenic plants, which are transformed by means of a DNA sequence according to the invention or a corresponding construct, respectively. Accordingly, there may be obtained e.g. transgenic plants, in which a GntI deficiency, for example on account of a missing or defective *GntI* gene or due to defects in the regulatory regions of this gene, has been removed by complementation using a construct derived from the cDNA sequences mentioned in the sequence protocol, wherein the expression of said construct is under the control of an active constitutive or inducible promoter, which may be additionally tissue specific. In this case, the GntI enzyme or protein expressed on account of the DNA of the invention contained in the construct and having GntI activity complements the GntI activity missing in the starting plant.

Also considered are transgenic plants, in which the GntI activity already present in the starting plant is increased by additional expression of the *GntI* transgene introduced by means of a construct according to the present invention. Up to now, the extremely low expression of the *GntI* gene *in vivo* accompanied by extremely low enzyme activity, which correspondingly was very difficult to detect, has been a main problem in the investigation of the enzyme N-acetyl glucosaminyl transferase I in plants. The problem of a too low GntI enzyme activity in plants may be overcome by the coexpression of a DNA according to the present invention.

In this case, it may be preferable for the transformation of plants to employ DNA according to the invention, additionally comprising a sequence region, which following expression enables a facilitated detection and/or enrichment and purification, respectively, of the protein product having GnTI activity. This is for example accomplished by the use of a specific DNA sequence for the expression of a recombinant GnTI enzyme, said sequence carrying a N-terminal or C-terminal sequence extension encoding an affinity marker. If it is additionally intended to provide an amino acid sequence portion between the GnTI enzyme and the affinity marker, which represents a recognition site for a specific protease, cleavage of the N-terminal or C-terminal sequence extension from the GnTI enzyme may be achieved by the subsequent use of this specific protease, and the GnTI enzyme thereby obtained in isolated form.

An example for this is the use of a DNA sequence according to the present invention, which codes for the recombinant GnTI enzyme with a C-terminal sequence extension, encoding the affinity marker AWRHPQFGG (*strep-tag*; ref. 39), and an intervening protease recognition site IEGR. The expression of the DNA according to the present invention provides GnTI enzymes with the C-terminal sequence extension mentioned, by means of which the expressed protein molecules specifically bind to a streptavidin derivatized matrix, and may thus be isolated. Then, by means of the protease factor Xa specifically recognizing the amino acid sequence IEGR, the GnTI portion of the protein molecules may be released. As an alternative, the complete protein may be removed from the streptavidin derivatized matrix by means of biotin or biotin derivatives.

A further example is represented by DNA sequences of the invention, encoding a protein which comprises multiple, e.g. 10, N-terminally added histidine residues (*His-tag*) in addition to a GnTI enzyme. Due to the N-terminal histidine residues, isola-

tion or purification, respectively, of the proteins expressed may be easily conducted by metal-chelate affinity chromatography (e.g. Ni sepharose) (cf. also Example 5).

5 Moreover, the invention comprises portions of such transgenic plants, adequately transformed plant cells, transgenic seeds and transgenic reproduction material.

10 A further important aspect of the invention is the use of the sequence information discussed above for the production of plants having reduced or lacking N-acetyl glucosaminyl transferase I activity.

15 The possibilities of identifying plants with reduced or lacking N-acetyl glucosaminyl transferase I activity due to a gene defect or a missing gene by means of antibodies of the invention or screening or hybridization probes of the invention have already been described above.

20 Two additional possibilities reside in the use according to the invention of antisense or sense constructs, respectively, which are derived from the DNA sequence of a plant *GntI* gene, for the production of transgenic plants with reduced or lacking N-acetyl glucosaminyl transferase I activity by means of
25 homology-dependent gene silencing (cf. ref. 16,17). The DNA sequence used as a starting sequence for the generation of the constructs, may be derived from the starting plant to be transformed itself but also from a different plant variety or
30 species. In particular, antisense or sense constructs as discussed under items iii) and iv) above are of use. Generally, the constructs employed comprise at least 50 to
35 200 and more base pairs.

40 In particular, the constructs employed for this purpose comprise at least 50 to 200 and more base pairs, with a sequence, which is derived on the basis of

- the cDNA sequences mentioned in the sequence protocol and/or the corresponding *GntI* genes and/or
- the derived DNA sequences discussed above and/or DNA fragments according to the present invention and/or
- the DNA sequences, in particular from other varieties and plant species, which encode N-acetyl glucosaminyl transferase I and which may be identified due to a hybridization under stringent conditions to hybridization or screening probes, as defined under items iii) and iv) above.

Generally, the constructs contain a strong constitutive or inducible promoter, which additionally may be tissue specific, by means of which the antisense or sense DNA sequence regions are controlled.

In the production of transgenic plants by integration of antisense construct(s) into the plant genome or by viral infection of starting plants or plant cells by means of virus containing antisense construct(s) for an extrachomosomal propagation and transcription of the antisense construct or the antisense constructs in infected plant tissue, it is intended to achieve a hybridization of *GntI*-gene transcripts to transcripts of the antisense DNA region at the RNA level, which prevents translation of the *GntI* mRNA. The result is a transgenic plant with strongly decreased contents of N-acetyl glucosaminyl transferase I, and thus, a strongly decreased corresponding enzyme activity.

For the transformation of plants according to the invention with antisense constructs, for example constructs may be employed, which hybridize to one of the complete cDNAs, mentioned in Fig. 2 and in the sequence protocol, or to corresponding regions thereof, generally comprising at least 50 to more than 200 base pairs. Moreover, particularly preferred is the use of fragments, the transcripts of which additionally cause a hybridization to a portion of the 5' untranslated region of the *GntI* mRNA, at which or in the proxi-

mity of which usually the attachment of ribosomes would occur. Examples of such constructs are shown in Fig. 4.

In view of the occurrence of an isoform in *Solanum tuberosum*, which probably is located in the cytoplasm due to lack of the membrane anchor (aa 10 to 29) of yet unknown function, it may be desirable to target only the N-acetyl glucosaminyl transferase I enzyme located in the Golgi cisternae, i.e. only that enzyme comprising the membrane anchor. One reason for this desire may be the effort or, in the individual case, also the requirement, to affect as little as possible the cytoplasmatic metabolism of the plant cell, for which the cytoplasmatic N-acetyl glucosaminyl transferase I possibly is of importance. For this purpose, antisense constructs may be used according to the present invention, which themselves or the transcripts of which, respectively, hybridize to a DNA or RNA region of the *GntI* gene or the *GntI* mRNA, comprising a part of the 5' untranslated region and the coding region including the membrane anchor. Generally, the extension of the region of hybridization up to position 266 of the cDNA in Fig. 2 and SEQ ID NO: 1 is considered harmless for the purpose mentioned above.

In the production of transgenic plants by integration of sense constructs into the plant genome or by viral infection of starting plants or plant cells by means of virus containing sense construct(s) for extrachromosomal propagation and expression of the construct or constructs in infected plant tissue, there are assumed hybridization phenomena in tobacco according to the work of Faske et al. (ref. 17), of said constructs to the endogenous *GntI* gene at a posttranscriptional or DNA level, respectively, which finally affect or prevent the translation of the *GntI* gene. Also in this case, the result are transgenic plants having reduced or even lacking N-acetyl glucosaminyl transferase I activity.

Methods for the stable integration of such antisense and sense constructs into the genome of plants, or for the viral infection of plants or plant cells, respectively, for an extrachromosomal propagation and transcription/expression of such constructs in infected plant tissue are well known to those skilled in the art. This includes the direct DNA transfer (e.g. into protoplasts by means of electroporation or by the addition of a high molecular osmotic agent as well as biolistic methods, by which DNA coated particles are shot into the plant tissue), such as the use of natural host/vector systems (e.g. agrobacteria or plant viruses). For viral infection of starting plants or plant cells by viruses containing appropriate constructs for extrachromosomal propagation and transcription/expression of the constructs in infected plant tissue, a variety of specific viruses, such as tobacco mosaic virus (TMV) or potato virus X, is available.

Representative plants, which are suitable for such integration, comprise dicotyledonous as well as monocotyledonous cultivated plants, in particular *Solanaceae* such as potato, tobacco, tomato and pepper. Additionally, banana, alfalfa, canola, beets, soybean, lettuce, corn, rice and grain, would be suitable target plants for the use of homologous antisense constructs. For example, the sequence from *Arabidopsis thaliana* mentioned in the sequence protocol appears to be particularly suitable as a starting sequence for the transformation according to the invention of *Brassicaceae*, such as canola plants, by means of sense or antisense constructs. Further plants of interest are any plants, which express glycoproteins of interest for medicine and research.

Generally, it should be noted, that the transformation according to the invention of plants, which in the corresponding region of the *GntI* gene exhibit a homology of $\geq 70\%$ at the nucleotide level to the employed antisense or sense constructs according to the present invention, typically results in

transgenic plants of the invention, which show the desired reduction of N-acetyl glucosaminyl transferase I activity.

Further, another possibility is seen in the targeted destruction (knock-out) of the N-acetyl glucosaminyl transferase I gene via gene targeting by means of homologous recombination (ref. 24) in a target plant using a suitable DNA fragment derived from the cDNA sequence of the present invention, similar to the procedure established for yeast systems and mammals.

Further, the present invention comprises transgenic plants, which have been transformed by the antisense or sense constructs mentioned above or the viruses containing the same, respectively, as well as parts of such transgenic plants, correspondingly transformed plant cells, transgenic seeds and transgenic reproduction material.

Methods of the production of transgenic plants, e.g. by means of agrobacteria- or virus-mediated as well as direct DNA transfer are known to those skilled in the art. Concerning representative plants for such a transformation, the above mentioned applies.

The plants of the invention and the plants obtained according to the invention, respectively, with reduced or lacking N-acetyl glucosaminyl transferase I activity, may be used according to the invention for the production of glycoproteins with minimal and uniform, i.e. defined, sugar residues. As discussed above, such glycoproteins are of great importance for medicine and research. As a reasonable source of raw material and food as well as due to their unproblematical disposal via composting, plants per se represent ideal bioreactors. According to the present invention, it is now possible to express biotechnologically or pharmaceutically relevant glycoproteins (e.g. therapeutics of low antigenic

potential for mammals) in cultivated plants, in which GnTI activity is highly reduced or completely absent.

Accordingly, the invention also comprises a method for the production of glycoproteins with minimal uniform and defined sugar residues, comprising cultivating a transgenic plant according to the invention, of parts of such plants or of plant cells transformed according to the invention, each expressing the desired glycoprotein, as well as isolating the desired glycoprotein from the cultivated material.

In this context, representative cultivated plants are Solanaceae, in particular potato, tobacco, tomato and pepper. Furthermore possible are banana, alfalfa, canola, beets, soybean, lettuce, corn, rice and grain.

The sequence of the enzymatically controlled and plant specific N-glycan modifications, which secretory glycoproteins are subjected to during passage through the Golgi apparatus of higher plants, is schematically shown in Fig. 1. The biosynthesis block due to lacking or insufficient N-acetyl glucosaminyl transferase I (GlcNAc transferase I) activity in a plant leads, instead of complex glycans, to the predominant formation of glycans of the $\text{Man}_5\text{GlcNAc}_2$ type, i.e. glycoproteins with uniform and well-defined sugar residues, which are of extremely high importance for medicine and research.

For this purpose, the genes encoding the desired glycoproteins may be expressed in their natural producing plants, which have been transformed according to the present invention e.g. by means of antisense or sense constructs to yield transgenic plants with reduced or missing N-acetyl glucosaminyl transferase I activity.

There is also the possibility to use transgenic plants of the invention displaying reduced or lacking N-acetyl glucosaminyl transferase I activity, which additionally have been transformed by the gene encoding the desired glycoprotein. In order to achieve this, constructs may be employed, which contain the gene encoding the desired glycoprotein under the control of a strong constitutive or inducible promoter, which is optionally tissue specific as well, and lead to the integration of the gene into the plant genome. Alternatively, the transformation may also be conducted by viral infection by means of a virus containing the gene for the desired glycoprotein for extrachromosomal propagation and expression of the gene. The glycoprotein may then be expressed in the respective host plant and obtained therefrom.

Naturally, as an alternative, the procedure may be such, that initially a transformation using an expression construct or virus containing the DNA encoding the glycoprotein is performed, and subsequently, another transformation with one or more of the antisense or sense constructs of the invention or with one or more viruses, containing the corresponding DNA, is performed. It is also possible to perform a simultaneous transformation using both constructs or using one virus containing the antisense or sense construct as well as the gene encoding the desired glycoprotein (piggyback version).

Within the scope of the present invention, there is also considered a viral overinfection of the transgenic plants according to the invention, in which integration of an antisense/sense construct and/or the gene encoding the desired glycoprotein into the genome has already occurred, by viruses containing the antisense/sense construct and/or the gene encoding the desired glycoprotein, for an additional extrachromosomal propagation and transcription or expression, respectively, of this DNA. As a result, the concentrations of antisense or sense DNA, respectively, or of the expressed glycoprotein may be increased in the transgenic plant cells.

It may prove to be practical for the production according to the invention of glycoproteins with defined glycosylation, to use tissue specific promoters in such cases, where it is intended to obtain the desired glycoproteins specifically only from certain parts of a plant such as tubers or roots. Today, for a large variety of plant tissues, tissue specific promoters are available, which drive expression of foreign genes specifically only in these tissues. By way of example, tuber specific promoters such as patatin class I (ref. 26) and proteinase inhibitor II promoters (ref. 27) may be mentioned. Under certain conditions, both promoters exhibit expression also in leaf tissue, i.e. they can be induced by high metabolite contents (for example sucrose) and in the case of the proteinase inhibitor II promoter also by mechanical lesion or by spraying with abscisic or jasmonic acid, respectively.

The use of tissue specific promoters may also be indicated in cases, where the DNA sequence or the transcription products or translation products thereof according to the invention, respectively, which are employed for the transformation, turn out to be detrimental to certain plant parts, e.g. due to a negative influence on the metabolism of the corresponding plant cells.

As a representative target glycoprotein, human glucocerebrosidase may be used for the therapy of the hereditary Gaucher's disease (ref. 25). In order to obtain human glucocerebrosidase (GC) with uniform and defined sugar residues, e.g. plants of the present invention which are transformed by means of antisense DNA, may be transformed with the gene encoding human glucocerebrosidase. For this purpose, the human glucocerebrosidase cDNA sequence (ref. 38) is modified at the 3' terminus by means of PCR using gene specific primers in a manner, that the recombinant enzyme carries a C-terminal sequence extension encoding an affinity marker (e.g.

AWRHPQFGG, strep-tag; ref. 39) and, optionally, also a protease recognition site (e.g. IEGR) between the GntI enzyme region and the affinity marker. The GC-cDNA sequence thus altered is expressed in *GntI* antisense plants of the present invention by using a strong and optionally tissue specific promoter (e.g. for potato under the control of the tuber specific B33 patatin promoter), so that the enzyme synthesized in these plants exclusively carries well defined N-glycans. The affinity marker is intended to facilitate the enrichment of the recombinant enzyme from the transgenic plants. In this case, the expressed protein molecules (GC-strep molecules) bind to a streptavidin derivatized matrix via the affinity marker sequence and can be released therefrom by means of biotin or biotin derivatives. The removal from the streptavidin derivatized matrix may also be carried out by means of catalytic amounts of a protease, which exhibits a specificity for the protease recognition site located between the GntI enzyme region and the affinity marker. In this case, only the GntI enzyme region is released from the matrix. This could be advantageous especially in that case, if the affinity marker sequence has a detrimental effect on the GntI activity.

Due to their terminal mannose residues, the $\text{Man}_5\text{GlcNAc}_2$ -glycans of the glucocerebrosidase obtained from the plants of the present invention will be recognized by macrophages as an uptake signal, and can thus directly be employed for the therapy of hereditary Gaucher's disease. Currently, a therapy is only possible upon expensive isolation and deglycosylation of native glucocerebrosidase (ref. 25).

Accordingly, the production of recombinant glycoproteins may be highly facilitated by the use of plant *GntI* sequences compared to conventional methods, e.g. the chemical deglycosylation of purified glycoproteins, which is technically demanding (ref. 25), or a difficult and expensive production in GntI deficient animal cell lines (ref. 7,10).

Description of the figures:

Fig. 1: Sequence of plant specific N-glycan modifications, which secretory glycoproteins are subject to during passage through the Golgi apparatus of higher plants (ref. 28). The biosynthesis block to complex modified glycans is based on a deficiency in GntI activity (which is either caused by a defective or missing GntI enzyme or by effective reduction of the *GntI* gene expression) and is indicated by a cross. Meaning of the symbols: (F) fucose residues, (X) xylose residues, (●) GlcNAc residues, (□) mannose residues.

Fig. 2: Full length cDNA sequence of a plant GntI from potato (*Solanum tuberosum* L.) and amino acid sequence deduced therefrom. By way of example, the complete cDNA of the membrane anchor containing *GntI* isoform from potato leaf tissue (A1) is illustrated. The EcoRI/NotI linkers at the 5' and 3' ends of the cDNA are highlighted by bold letters, the binding sites of the degenerate oligonucleotides used for obtaining the RT-PCR probe are underlined. In contrast to already published animal GntI sequences, the protein sequence derived from the potato cDNA clones contains a potential N-glycosylation site: Asn-X(without Pro)-Ser/Thr, which is indicated by an asterisk. The region of the membrane anchor is highlighted in italics (aa 10 to 29). The start of the isoform (A8), which is potentially located in the cytosol, is indicated by an arrow.

Fig. 3: A, Degree of identity or similarity, respectively, of the amino acid sequence deduced from a complete *GntI* cDNA sequence from potato (A1) in comparison to other GntI sequences of animal organisms, which have been selected from data bases. Identical amino acid posi-

tions (in %) are printed in bold letters, similar amino acid positions are given in brackets underneath. Meaning of the abbreviations: Hu, human; Ra, rat; Mo, mouse; Ce, *Caenorhabditis elegans* (round-worm); St, *Solanum tuberosum* (potato).

B, Comparison of the derived amino acid sequences of different plant *GntI*-cDNA clones. A_Stb-A1, GntI from potato leaf; B_Ntb-A9, GntI from tobacco leaf (A9); C_Atb-Full, GntI from *Arabidopsis thaliana*. Identical aa are highlighted in black, similar aa in light grey.

Fig. 4: Cloning strategy of the *GntI*-antisense constructs used. Following fill-in of the ends, a NotI linker was introduced into the SalI restriction site of the polylinker region of the plant expression vector pA35 (=pA35N) (ref. 29), and the complete A1-*GntI*-cDNA was inserted into pA35N via NotI. The corresponding antisense construct (=pA35N-Alas) was inserted into binary vector pBin19 (ref. 30) via EcoRI and HindIII. Additionally, following PCR amplification, a 5' fragment of the A1-*GntI*-cDNA comprising 270 bp was cloned into pA35N via XbaI and NotI restriction sites in antisense orientation (=pA35N-A1-short) and also inserted into pBin19. Abbreviations; Numerals in brackets, positions of the restriction sites in the A1-*GntI*-cDNA (in base pairs); pBSK, cloning vector (Stratagene); pGEM3Z, cloning vector (Promega); CaMVp35S, constitutive 35S promoter of cauliflower mosaic virus; OCSpA, polyadenylation signal of octopine synthase; pNOS, promoter of nopaline synthase; NEO, neomycin phosphotransferase (selection marker, confers kanamycin resistance); NOSpA, polyadenylation signal of nopaline synthase; LB/RB, left/right border of the T-DNA of the binary vector; arrow, translation initiation (ATG); A8, start of the GntI isoform,

which is potentially located in the cytosol (7 aa substitutions in comparison to A1).

Fig. 5: Extent of suppression of complex glycoprotein modification in transgenic potato plants transformed with the long *GntI* antisense construct (cf. Fig. 4). A, Coomassie-stained SDS gel from leaf extracts; B, Western-blot analysis (Ref. 13,33) of parallel samples developed with a complex-glycan antiserum (Ref. 12,13). The lanes contain 30 µg each of total protein: *cgl*(Ara), Arabidopsis *cgl* mutant (Ref. 13); WT(Desi), wild-type potato; the numerals refer to individual transgenic potato plants; the arrows represent molecular weight standards of 66, 45, 36 and 29 kDa, respectively.

Fig. 6: Detection of specificity of the generated GntI antiserum following cell fractionation (Ref. 40) of tobacco callus material. For Western-blot analysis (Ref. 13,33) 30 µg of protein were applied per lane. The antiserum was used in 1:1000 dilution. Lane 1, homogenate following separation of cellular debris; lane 2, vesicle fraction following column chromatography; lane 3, sucrose gradient fraction I (microsomes); lane 4, sucrose gradient fraction II (plastids); lane 5, antigen used for immunization (recombinant GntI fusion protein); arrow, molecular weight of about 49 kDa.

Explanation of the abbreviations used in the text:

Aa, amino acid(s); bp, base pair(s); EMS, Ethyl methane sulfonate (mutagenic agent); F2, second filial generation; Fuc, fucose; Glc, glucose; GlcNAc, N-acetyl glucosamine; GntI, N-acetyl glucosaminyl transferase I (EC 2.4.1.101); *GntI*, gene for GntI (nuclear encoded); kDa, kilodalton; Man, mannose;

PCR, polymerase chain reaction; PAGE, polyacrylamide gel electrophoresis; ref., reference; RT-PCR, reverse transcription coupled polymerase chain reaction; SDS, sodium dodecyl sulfate; var., variety; Xyl, xylose.

In the following, the invention will be described in more detail by means of examples, which are only intended to illustrate the invention and shall not limit the invention in any manner.

Example 1

Isolation and characterization of plant *GntI* cDNA clones.

Total RNA was isolated from potato and tobacco leaf tissue, and cDNA fragments of about 90 bp were amplified by means of RT-PCR in combination with degenerate primers (procedure analogous to ref. 31), which were derived from conserved amino acid regions of known GntI sequences from animal organisms (sense primer 1*, 5'-TG(CT) G(CT)I (AT)(GC)I GCI TGG (AC)A(CT) GA(CT) AA(CT)-3'; antisense primer 3*, 5'-CCA ICC IT(AG) ICC (ACGT)G(CG) (AG)AA (AG)AA (AG)TC-3'; 30 pmol of each primer per 50 µl PCR assay at an annealing temperature of 55°C and 45 cycles). Following gel elution, the ends of the PCR products were repaired (i.e. blunt ended using DNA polymerase I and phosphorylated using T4 polynucleotide kinase) and cloned into the EcoRV restriction site of pBSK (Stratagene). By comparison with known GntI sequences between the primers (arrows), the identity of the derived amino acid sequences from the potato and tobacco RT-PCR products could be confirmed as being homologous; ⇒Q(R/M)QFVQDP(D/Y)ALYRS⇐ (homologous aa are underlined). Of one clone each, radio-labelled probes were synthesized by means of PCR (standard PCR assay using degenerate primers as above, nucleotide mixture without dCTP, but instead with 50 µCi α-³²P-dCTP [>3000 Ci/mMol]), and different cDNA libraries were screened for *GntI* containing clones using the corresponding homologous potato or tobacco probes, respectively (procedure analogous

to ref. 31; the stringent hybridization conditions have already been described in the text above). The cDNA libraries were prepared from mRNA of young and still growing plant parts (sink tissues). Following cDNA synthesis and ligating EcoRI/NotI adaptors (cDNA synthesis kit, Pharmacia) EcoRI compatible lambda arms were ligated, those packaged and used to transfect *E. coli* XL1 Blue cells (Lambda ZAPII cloning and packaging system, Stratagene). Following amplification of the libraries, one full-length *GntI* clone each was isolated from a potato leaf sink library (A1 according to Fig. 2 and SEQ ID NO: 1) and a tobacco leaf sink library (A9 according to SEQ ID NO: 3), as well as two additional clones from a tuber sink library (A6, A8). The deduced *GntI* amino-acid sequences contain a potential N-glycosylation site, Asn-X(without Pro)-Ser/Thr, in contrast to those of animals. One of the tuber *GntI* cDNA sequences carries stop codons in all three reading frames in front of the first methionine (A8). The coding region shows high homology to the longer tuber clone (A6) (only 2 aa substitutions), but displays a completely different 5' non-translated region. Furthermore, the membrane anchor characteristic for the Golgi enzyme is missing, so that this *GntI* isoform might be located in the cytosol. Sequence comparisons carried out by means of the gap or pileup option, respectively, and the box option of the gcg software package (J. Devereux, P. Haeberli, O. Smithies (1984) Nucl. Acids Res. 12: 387-395) indicate, that the deduced plant *GntI* amino-acid sequences exhibit only 30-40% identity and 57-59% similarity to those of animal organisms (Fig. 3A), while they are highly homologous among each other (75 - 90% identity, Fig. 3B).

The procedure in the case of *Arabidopsis thaliana* was analogous, wherein for the preparation of a specific probe first a partial *GntI* sequence was amplified by RT-PCR using *GntI* sense primer 4A (5'-ATCGGAAAGCTT**GGATCC** CCA GTG GC(AG) GCT GTA GTT GTT ATG GCT TGC-3'; HindIII restriction site underlined, BamHI printed in bold) and antisense primer 3*, as defined

above. First, a 5'-incomplete cDNA clone was isolated from a phage library (Lambda Uni-Zap) using this probe. By means of a vector insert PCR, the missing 5'-terminus was amplified from another library (via an unique SpeI restriction site in the 5' region) and assembled to yield a full-length cDNA sequence. The nucleic acid sequence determined by means of sequencing is listed in SEQ ID NO: 5.

Example 2

Functional complementation of a GntI defect using *GntI* cDNA upon transient expression in protoplasts of the *Arabidopsis thaliana* *cgl* mutant.

Approximately 4 weeks subsequent to sowing, protoplasts were isolated from leaves of *cgl* mutants cultivated under sterile conditions (nonstainer plants following 5 backcrosses, ref. 13), transformed with expression constructs of the complete *GntI* cDNA sequences (NotI cDNA fragments, cf. Fig. 4) in sense (pA35N-A1s or pA35N-A9s, respectively) or antisense orientation (pA35N-A1as or pA35N-A9as, respectively), and cultivated for 96 h at room temperature in the dark (50 µg of plasmid DNA each per 1 million protoplasts, PEG method according to ref. 32). Subsequent SDS-PAGE of the protoplast extracts and Western-blot analysis (analogous to ref. 13, 33) indicated functional complementation of the GntI defect, i.e. complex glycosylation of numerous protein bands upon transient expression of the potato A1 and tobacco A9 sense constructs, but not of the corresponding antisense constructs in protoplasts of the *Arabidopsis cgl* mutant (data not shown).

Example 3

Cloning of the binary expression constructs
pBin-35-Alas and pBin-35-Al-short (cf. Fig. 4).

5 Into the SalI restriction site of the polylinker region
(corresponding to the one of pUC18) of plant expression
vector pA35 (ref. 29), a NotI linker was introduced subse-
quently to the fill-in of the ends (=pA35N), and the complete
10 Al-*GntI*-cDNA (nucleotides 9 to 1657; according to the cDNA in
Fig. 2) was inserted into pA35N via NotI (sense construct
pA35N-Als and antisense construct pA35N-Alas, respectively).
The expression cassettes of the sense and antisense
15 constructs, respectively, were isolated via the terminal
restriction sites (filled-in NcoI restriction site, partial
post digestion with HindIII) as a fragment of about 2410 bp
and inserted into the EcoRI (filled-in) and HindIII
20 restriction sites of the binary vector pBin19 (Ref. 30)
(=pBin-35-Als and pBin-35-Alas, respectively). The EcoRI
restriction site of the vector is restored by fusion with the
equally filled-in NcoI restriction site of the fragment. By
25 means of a standard PCR assay (sense primer: KS sequencing
primer (Stratagene) extended for PCR, 5'-GGC CCC CCC TCG AGG
TCG ACG GTA TCG-3'; antisense primer: 5'-GGGCCTCTAGACTCGAG
AGC (CT)AC TAC TCT TCC TTG CTG CTG GCT AAT CTT G-3', XbaI
30 restriction site underlined, XhoI restriction site in ita-
lics), there was additionally amplified a 5'-fragment of the
GntI cDNA at an annealing temperature of 50°C (nucleotides 9
to 261, according to the cDNA in Fig. 2 and SEQ ID NO: 1).
35 The PCR product was digested with XbaI (within the antisense
primer) and NotI (within the 5'-linker of the cDNA), isolated
as a fragment of about 260 bp and cloned into pA35N (=pA35N-
Al-short). The expression cassette of the short antisense
40 construct was also inserted into pBin19 (=pBin-35-Al-short)
as a EcoRI/HindIII fragment (about 1020 bp).

Example 4

Transformation of agrobacteria by means of the binary *GntI* constructs and regeneration of transgenic potato and tobacco plants, respectively, from infected leaf discs.

5 The binary antisense *GntI* constructs (pBin-35-Alas and pBin-35-A1-short) were transformed into the Agrobacterium strain GV2260 (ref. 34, 35). By way of example, sterile leaf discs of potato plants var. Désirée and of tobacco plants var. Wisconsin 38 were infected with the recombinant agrobacterial lines (50 µl of a fresh overnight culture in 10 ml liquid 2MS medium: 2% sucrose in Murashige & Skoog salt/vitamin standard medium, pH 5.6; small pieces of leaf without midrip; co-cultivation for 2 days in the dark in phytotrons). Subsequent to washing of the infected leaf pieces in 2MS medium with 250 µg/ml claforan, transgenic plants were regenerated from said pieces in tissue culture under kanamycin selection (potato protocol ref. 26; tobacco protocol ref. 36) and analyzed for reduced GntI activity (exemplary shown in Fig. 5 for transgenic potato plants). As apparent from Fig. 5, antisense suppression of complex glycoprotein modification was successful in transgenic potato plant #439. The determined reduction of complex glycoprotein modification was stable in this transformant over the entire investigation period of several months and has been verified in three tests which were performed in an interval of about 1 month each. For the respective transgenic tobacco plants, analogous results were obtained.

Example 5

Production of recombinant potato GntI protein
(for the production of antibodies).

40 Recombinant GntI carrying 10 additional N-terminal histidine residues (His-tag) was produced in *E. coli* by means of the pET system (Novagen) and purified by metal-chelate affinity chromatography. A cDNA fragment comprising nucleotides 275-

1395 of the potato *GntI* cDNA (corresp. to aa 75-446, Fig. 2 and SEQ ID NO: 1 and 2, respectively) was amplified by standard PCR (annealing temperature of 50°C, 30 cycles, ref. 31) (sense primer *GntI*-5'fus: 5'-CATGGATCC CTC GAG AAG CGT CAG GAC CAG GAG TGC CGG C-3'; antisense primer *GntI*-3'stop: 5'-ATCCCGGGATCCG CTA CGT ATC TTC AAC TCC AAG TTG-3'; XhoI and BamHI restriction sites, respectively, are underlined, stop codon in italics), and inserted into vector pET16b (Novagen) (=pET-His-A1) via the restriction sites of the synthetic primer (5'-XhoI-*GntI*-BamHI-3'). Following propagation and analysis in *E. coli* XL1-Blue (Stratagene) the construct was stored as a glycerol culture. Competent *E. coli* BL21(DE3) pLysS cells (Novagen) were transformed with pET-His-A1 for overexpression. Addition of IPTG (Isopropyl-1-thio-β-D-galactopyranoside, at 0.5-2 mM) to a BL21 culture in logarithmic growth phase, initially induces the expression of T7 RNA polymerase (from the bacterial chromosome), and thus, also the expression of the recombinant fusion protein under control of the T7 promoter in pET vectors (Novagen). By means of metal-chelate chromatography using TALON matrix (Clontech), recombinant potato GntI was purified from induced BL21:pET-His-A1 cells under denaturing conditions via its His-tag (manufacturer's protocol, Novagen), and the preparation was verified with respect to homogeneity by means of SDS-PAGE.

Example 6

Raising of polyclonal antibodies in rabbits.

Recombinant potato GntI (from Expl. 5) was used as an antigen. Following the harvest of some milliliters of pre-immune serum, the rabbits were subcutaneously injected with 300-500 µg of affinity-purified protein together with 25 µg of GMDP adjuvant (Gerbu) in intervals of three weeks. Subsequent to three basis injections, the animals were bled from the ear vein 12 to 14 days after the respective successive injection (boost), the serum harvested (ref. 37) and tested for recognition of

recombinant GnTI by Western-blot analyses (dilution 1:200 to 1:2000). The antiserum of the boosts resulting in the lowest background-to-signal ratio were mixed with 0.04% sodium azide, aliquoted and kept at +4°C or for long-term storage at -20°C, respectively. As shown in Fig. 6, Western-blot analyses of tobacco callus cells (BY-2 suspension culture) revealed a specific GnTI signal in enriched microsomal fractions, which indicates, that antibodies raised against the recombinant protein specifically recognize plant GnTI. The detection was carried out with enriched microsomal fractions (ER and Golgi vesicles), since - due to low amounts - it is not possible to detect GnTI protein in crude plant extracts by means of the employed Western-blot method.

References:

- 1) R Kornfeld, S Kornfeld (1985) Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 54: 631-664
- 2) GP Kaushal, T Szumilo, AD Elbein (1988). Structure and biosynthesis of plant N-linked glycans. In J Preiss (editor) *The Biochemistry of Plants*, Vol 14: Carbohydrates. Academic Press, San Diego, CA, pp 421-463
- 3) L Faye, MJ Chrispeels (1989) Apparent inhibition of β -fructosidase secretion by tunicamycin may be explained by breakdown of the unglycosylated protein during secretion. *Plant Physiol* 89: 845-851
- 4) TW Rademacher, RB Parekh, RA Dwek (1988) Glycobiology. *Annu Rev Biochem* 57: 785-838
- 5) A Sturm (1991) Heterogeneity of the complex N-linked oligosaccharides at specific glycosylation sites of 2 secreted carrot glycoproteins. *Eur J Biochem* 199: 169-179
- 6) K Olden, BA Bernard, MJ Humphries, T Yeo, SL White, SA Newton, HC Bower, JB Parent (1985) Function of glycoprotein glycans. *Trends Biochem Sci* 10: 78-82
- 7) P Stanley (1989) Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol Cell Biol* 9: 377-383
- 8) R Kumar, J Yang, RD Larsen, P Stanley (1990) Cloning and expression of N-acetylglucosaminyltransferase I, the medial Golgi transferase that initiates complex N-linked

carbohydrate formation. Proc Natl Acad Sci USA 87: 9948-9952

- 5 9) JW Dennis, S Laferte, C Waghorne, ML Breitman, RS Kerbel
(1987) $\beta \rightarrow 6$ branching of Asn-linked oligosaccharides is
directly associated with metastasis. Science 236: 582-585
- 10 10) MN Fukuda (1990) HEMPAS disease: genetic defect of glyco-
sylation. Glycobiology 1: 9-15
- 15 11) MN Fukuda, KA Masri, A Dell, L Luzzatto, KW Moremen
(1990) Incomplete synthesis of N-glycans in congenital
dyserythropoietic anemia type II caused by a defect in the
gene encoding α -mannosidase II. Proc Natl Acad Sci USA
87: 7443-7447
- 20 12) M Laurière, C Laurière, MJ Chrispeels, KD Johnson, A
Sturm (1989) Characterization of a xylose-specific anti-
serum that reacts with the complex asparagine-linked gly-
25 cans of extracellular and vacuolar glycoproteins. Plant
Physiol 90: 1182-1188
- 30 13) A von Schaewen, A Sturm, J O'Neill, MJ Chrispeels (1993)
Isolation of a mutant *Arabidopsis* plant that lacks
N-acetyl glucosaminyl transferase I and is unable to syn-
thesize Golgi-modified complex N-linked glycans. Plant
35 Physiol 102: 1109-1118
- 40 14) JK-C Ma, MB Hein (1995) Plant antibodies for immuno-
therapy. Plant Physiol 109: 341-346
- 15) AS Moffat (1995) Medical applications: Exploring trans-
genic plants as a new vaccine source. Science 268: 658-

660 (summary of two original publications in the same issue)

- 16) CB Taylor (1997) Comprehending cosuppression. Plant Cell
9: 1245-1249 (summary of several original publications in
the same issue)
- 17) M Faske, JE Backhausen, M Sendker, M Singer-Bayrle, R
Scheibe, A von Schaewen (1997) Transgenic tobacco plants
expressing pea chloroplast *Nmdh* cDNA in sense and anti-
sense orientation: Effects on NADP-MDH level, stability
of transformants, and plant growth. Plant Physiol 115:
705-715
- 18) R Koes, E Souer, A van Houwelingen, L Mur, C Spelt, F
Quattrocchio, J Wing, B Oppedijk, S Ahmed, T Maes, T
Gerats, P Hoogeveen, M Meesters, D Kloos, JNM Mol (1995)
Targeted gene inactivation in petunia by PCR-based selec-
tion of transposon insertion mutants. Proc Acad Sci USA
92: 8149-8153
- 19) EC McKinney, N Ali, A Traut, KA Feldmann, DA Belostotsky,
JM McDowell, RB Meagher (1995) Sequence-based identifica-
tion of T-DNA insertion mutations in *Arabidopsis*: actin
mutants *act2-1* and *act4-1*. Plant J 8: 613-622
- 20) F Altmann, G Kornfeld, T Dalik, E Staudacher, J Glössl
(1993) Processing of asparagine-linked oligosaccharides
in insect cells. N-acetylglucosaminyl transferase I and
II activities in cultured lepidopteran cells. Glycobiol-
ogy 3: 619-625
- 21) A Sturm, KD Johnson, T Szumilo, AD Elbein, MJ Chrispeels
(1987) Subcellular localization of glycosidases and gly-

cosyltransferases involved in the processing of N-linked oligosaccharides. *Plant Physiol* 85: 741-745

- 22) GM Church, W Gilbert (1984) Genomic sequencing. *Proc Acad Sci USA* 81: 1991-1995
- 23) J Sambrook, EF Fritsch, T Maniatis (1989) *Molecular cloning: a laboratory manual* (2nd edn), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- 24) H Puchta, B Hohn (1996) From centiMorgans to base pairs: homologous recombination in plants. *Plant Sci* 1: 340-348
- 25) NW Barton, FS Furbish, GJ Murray, M Garfield, RO Brady (1990) Therapeutic response to intravenous infusions of glucocerebrosidase in a patient with Gaucher disease. *Proc Natl Acad Sci USA* 87: 1913-1916
- 26) M Rocha-Sosa, U Sonnewald, W-B Frommer, M Stratmann, J Schell, L Willmitzer (1989) Both developmental and metabolic signals activate the promoter of a class I patatin gene. *EMBO J* 8: 23-29
- 27) T Hildmann, M Ebner, H Pena-Cortes, JJ Sanchez-Serrano, L Willmitzer, S Prat (1992) General roles of abscisic and jasmonic acids in gene activation as a result of mechanical wounding. *Plant Cell* 4: 1157-1170
- 28) KD Johnson, MJ Chrispeels (1987) Substrate specificities of N-acetylglucosaminyl-, fucosyl-, and xylosyltransferases that modify glycoproteins in the Golgi apparatus of bean cotyledons. *Plant Physiol* 84: 1301-1308
- 29) H Höfte, L Faye, C Dickinson, EM Herman, MJ Chrispeels (1991) The protein-body proteins phytohemagglutinin and

tonoplast intrinsic protein are targeted to vacuoles in leaves of transgenic tobacco. *Planta* 184: 431-437

- 5 30) M Bevan (1984) Binary *Agrobacterium* vectors for plant transformation. *Nucl Acids Res* 12: 8711-8721
- 10 31) K Graeve, A von Schaewen, R Scheibe (1994) Purification, characterization and cDNA sequence of glucose-6-phosphate dehydrogenase from potato (*Solanum tuberosum* L.). *Plant J* 5: 353-361
- 15 32) B Damm, R Schmidt, L Willmitzer (1989) Efficient transformation of *Arabidopsis thaliana* using direct gene transfer to protoplasts. *Mol Gen Genet* 213: 15-20
- 20 33) A von Schaewen, M Stitt, R Schmidt, L Willmitzer (1990) Expression of a yeast-derived invertase in the cell wall of tobacco and *Arabidopsis* plants leads to accumulation of carbohydrate, inhibition of photosynthesis and strongly influences growth and phenotype of transgenic tobacco
- 25 plants. *EMBO J.* 9: 3033-3044
- 30 34) R Deblaere, B Bytebier, H De Greve, F Debroeck, J Schell, M van Montagu, J Leemans (1985) Efficient octopine Ti plasmid-derived vectors for *Agrobacterium* mediated gene transfer to plants. *Nucl Acids Res* 13: 4777-4788
- 35 35) R Höfgen, L Willmitzer (1988) Storage of competent cells for *Agrobacterium* transformation. *Nucl Acids Res* 16: 9877
- 40 36) T Voelker, A Sturm, MJ Chrispeels (1987) Differences in expression between two seed lectin alleles obtained from normal and lectin-deficient beans are maintained in transgenic tobacco. *EMBO J* 6: 3571-3577

- 37) E Harlow, D Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- 5 38) J Sorge, C West, B Westwood, E Beutler (1985) Molecular cloning and nucleotide sequence of human cerebrosidase cDNA. Proc Natl Acad Sci USA 82: 7289-7293
- 10 39) TGM Schmidt, A Skerra (1993) The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. Prot Engineering 6: 109-122
- 15 40) M van der Wilden, NR Gilkes, MJ Chrispeels (1980) The endoplasmic reticulum of mung bean cotyledons: role in the accumulation of hydrolases in protein bodies during seedling growth. Plant Physiol. 66: 390-394
- 20
- 25
- 30
- 35
- 40

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: von Schaewen, Antje Dr. rer. nat.
- (B) STREET: Natruperstrasse 169a
- (C) CITY: Osnabrueck
- (E) COUNTRY: Germany
- (F) POSTAL CODE (ZIP): D-49076
- (G) TELEPHONE: +49-541-684029

(ii) TITLE OF INVENTION: Plant gntI sequences and the use thereof for the production of plants having reduced or lacking N-acetyl glucosaminyl transferase I (GnTI) activity

(iii) NUMBER OF SEQUENCES: 6

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1669 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Solanum tuberosum
- (B) STRAIN: Desiree
- (D) DEVELOPMENTAL STAGE: Sink organ
- (F) TISSUE TYPE: Mesophyll
- (G) CELL TYPE: Leaf cells

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: Lambda ZAP II (Eco RI)
- (B) CLONE: gntI-A1(K)

(ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 659..667
- (D) OTHER INFORMATION: /function= "Asn codon in this context is a potential glycosylation site"
/product= "N-glycosylation consensus sequence"
/phenotype= "N-glycans modulate protein properties"
/standard_name= "N-glycosylation site"
/label= pot-CHO
/note= "GnTI-coding sequences from animals do not contain this feature"

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION:53..1393
 (C) IDENTIFICATION METHOD: experimental
 (D) OTHER INFORMATION:/codon_start= 53
 /function= "initiates complex N-glycans on
 secretory glycoproteins"
 /EC_number= 2.4.1.101
 /product=
 "beta-1,2-N-acetylglucosaminyltransferase I"
 /evidence= EXPERIMENTAL
 /gene= "cgl"
 /standard_name= "gntI"
 /label= ORF
 /note= "first gntI sequence from potato (unpublished)"

(ix) FEATURE:
 (A) NAME/KEY: 5'UTR
 (B) LOCATION:15..52

(ix) FEATURE:
 (A) NAME/KEY: 3'UTR
 (B) LOCATION:1394..1655

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION:80..139
 (D) OTHER INFORMATION:/function= "membrane anchor (amino
 acids 10-29)"
 /product= "hydrophobic amino acid stretch in GntI"
 /standard_name= "membrane anchor of a type II
 Golgi protein"
 /note= "identified by comparison with GntI sequences
 from animals"

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:1..14
 (D) OTHER INFORMATION:/function= "used for cloning the
 cDNA library in Lambda ZAPII"
 /product= "EcoRI/NotI-cDNA adapter"
 /number= 1

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:1656..1669
 (D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
 /number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAATTCGCGG CCGCCTGAGA AACCTCGAA TTCAATTTTCG CATTTGGCAG AG ATG	55
Met	
1	
AGA GGG AAC AAG TTT TGC TTT GAT TTA CGG TAC CTT CTC GTC GTG GCT	103
Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val Ala	
5 10 15	
GCT CTC GCC TTC ATC TAC ATA CAG ATG CGG CTT TTC GCG ACA CAG TCA	151
Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln Ser	

20	25	30	
GAA TAT GTA GAC CGC CTT GCT GCT GCA ATT GAA GCA GAA AAT CAT TGT Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His Cys 35 40 45			199
ACA AGT CAG ACC AGA TTG CTT ATT GAC AAG ATT AGC CAG CAG CAA GGA Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gln Gly 50 55 60 65			247
AGA GTA GTA GCT CTT GAA GAA CAA ATG AAG CAT CAG GAC CAG GAG TGC Arg Val Val Ala Leu Glu Glu Gln Met Lys His Gln Asp Gln Glu Cys 70 75 80			295
CGG CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG GGC ATA AAA Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile Lys 85 90 95			343
AAG TTA ATC GGA GAT GTG CAG ATG CCA GTG GCA GCT GTA GTT GTT ATG Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val Met 100 105 110			391
GCT TGC AGT CGT ACT GAC TAC CTG GAG AGG ACT ATT AAA TCC ATC TTA Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile Leu 115 120 125			439
AAA TAC CAA ACA TCT GTT GCA TCA AAA TAT CCT CTT TTC ATA TCC CAG Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln 130 135 140 145			487
GAT GGA TCA AAT CCT GAT GTA AGA AAG CTT GCT TTG AGC TAT GGT CAG Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly Gln 150 155 160			535
CTG ACG TAT ATG CAG CAC TTG GAT TAT GAA CCT GTG CAT ACT GAA AGA Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu Arg 165 170 175			583
CCA GGG GAA CTG GTT GCA TAC TAC AAG ATT GCA CGT CAT TAC AAG TGG Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp 180 185 190			631
GCA TTG GAT CAG CTG TTT CAC AAG CAT AAT TTT AGC CGT GTT ATC ATA Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile Ile 195 200 205			679
CTA GAA GAT GAT ATG GAA ATT GCT GCT GAT TTT TTT GAC TAT TTT GAG Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe Glu 210 215 220 225			727
GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG GCT ATT TCT Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile Ser 230 235 240			775
TCT TGG AAT GAC AAT GGA CAA AGG CAG TTC GTC CAA GAT CCT GAT GCT Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp Ala 245 250 255			823

000000-000000

CTT TAC CGC TCA GAC TTT TTT CCT GGT CTT GGA TGG ATG CTT TCA AAA Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser Lys 260 265 270	871
TCA ACT TGG TCC GAA CTA TCT CCA AAG TGG CCA AAG GCT TAC TGG GAT Ser Thr Trp Ser Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp 275 280 285	919
GAC TGG CTA AGG CTG AAA GAA AAT CAC AGA GGT CGA CAA TTT ATT CGC Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile Arg 290 295 300 305	967
CCA GAA GTT TGC AGA ACG TAC AAT TTT GGT GAG CAT GGT TCT AGT TTG Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu 310 315 320	1015
GGG CAG TTT TTT AAG CAG TAT CTT GAG CCA ATT AAG CTA AAT GAT GTC Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val 325 330 335	1063
CAG GTT GAT TGG AAG TCA ATG GAC CTA AGT TAC CTT TTG GAG GAC AAC Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp Asn 340 345 350	1111
TAT GTG AAA CAC TTT GGC GAC TTG GTT AAA AAG GCT AAG CCC ATC CAC Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile His 355 360 365	1159
GGA GCT GAT GCT GTT TTG AAA GCA TTT AAC ATA GAT GGT GAT GTG CGT Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val Arg 370 375 380 385	1207
ATT CAG TAC AGA GAC CAA CTA GAC TTT GAA GAT ATC GCT CGA CAG TTT Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln Phe 390 395 400	1255
GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGG GCA GCA TAT AAA Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr Lys 405 410 415	1303
GGG ATA GTA GTT TTC CGG TTT CAA ACA TCT AGA CGT GTG TTC CTT GTT Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu Val 420 425 430	1351
TCC CCT GAT TCT CTT CGA CAA CTT GGA GTT GAA GAT ACT TAG Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr * 435 440 445	1393
CGAAGATATG ATTGGAGCCT GAGCAACAAT TTAGACTTAT TTGGTAGGAT ACATTTGAAA	1453
GAGCTGACAC GAAAAGTATG ACTACCAGTA GCTACATGCA ACATTTTAAT GTTAATGGAA	1513
GGAACCCACT GCTTATTGTT GGAATGGATG AATCATCACC ACATCCTATT ATTCAAGTTT	1573
ACAAACATAA AGAGGAAATG TTGCCCTATA AAAACAAATT TTTTGTCTCT AAGAAGGAAC	1633
GTTACGATTA TGAGCAACTT TGGCGGCCGC GAATTC	1669

000000-999999

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 447 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

Met Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val
 1           5           10           15
Ala Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln
          20           25           30
Ser Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His
          35           40           45
Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gln
          50           55           60
Gly Arg Val Val Ala Leu Glu Glu Gln Met Lys His Gln Asp Gln Glu
          65           70           75           80
Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile
          85           90           95
Lys Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val
          100          105          110
Met Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile
          115          120          125
Leu Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser
          130          135          140
Gln Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly
          145          150          155          160
Gln Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu
          165          170          175
Arg Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys
          180          185          190
Trp Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile
          195          200          205
Ile Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe
          210          215          220
Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile
          225          230          235          240
Ser Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp
          245          250          255
Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser
          260          265          270

```

Lys Ser Thr Trp Ser Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp
 275 280 285

Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile
 290 295 300

Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser
 305 310 315 320

Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp
 325 330 335

Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp
 340 345 350

Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile
 355 360 365

His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val
 370 375 380

Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln
 385 390 395 400

Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr
 405 410 415

Lys Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu
 420 425 430

Val Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr *
 435 440 445

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1737 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Nicotiana tabacum
- (B) STRAIN: Samsun NN
- (D) DEVELOPMENTAL STAGE: Sink organ
- (F) TISSUE TYPE: Mesophyll
- (G) CELL TYPE: Leaf cells

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: Lambda ZAP II (Eco RI)
- (B) CLONE: gntI-A9(T)

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:733..741
- (D) OTHER INFORMATION:/function= "Asn codon in this context is a potential glycosylation site"
/product= "N-glycosylation consensus sequence"
/phenotype= "N-glycans modulate protein properties"
/standard_name= "N-glycosylation site"
/label= pot-CHO
/note= "GnTI sequences from animals do not contain this feature"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:127..1467
- (C) IDENTIFICATION METHOD: experimental
- (D) OTHER INFORMATION:/codon_start= 127
/function= "initiates complex N-glycans on secretory glycoproteins"
/EC number= 2.4.1.101
/product= "beta-1,2-N-acetylglucosaminyltransferase I"
/evidence= EXPERIMENTAL
/gene= "cgl"
/standard_name= "gntI"
/label= ORF
/note= "first gntI sequence from tobacco (unpublished)"

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION:15..126

(ix) FEATURE:

- (A) NAME/KEY: 3'UTR
- (B) LOCATION:1468..1723

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:154..213
- (D) OTHER INFORMATION:/function= "membrane anchor (amino acids 10-29)"
/product= "hydrophobic amino acid stretch in GnTI"
/standard_name= "membrane anchor of a type II golgi protein"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1..14
- (D) OTHER INFORMATION:/function= "use for cloning the cDNA library in Lambda ZAPII"
/product= "EcoRI/NotI-cDNA adapter"
/number= 1

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1724..1737
- (D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GAATTCGCGG CCGCCATTGA CTTGATCCTA ACTGAACAGG CAAAGTAAAT CCAGCGATGA	60
AACACTCATA ACTGAACACT GAGAGACTAT TCGCTTTCTC CTAAAGCCTT CAATCGAATT	120
CGCACG ATG AGA GGG AAC AAG TTT TGC TGT GAT TTC CGG TAC CTC CTC	168
Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu	
450 455 460	
ATC TTG GCT GCT GTC GCC TTC ATC TAC ACA CAG ATG CGG CTT TTT GCG	216
Ile Leu Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala	
465 470 475	
ACA CAG TCA GAA TAT GCA GAT CGC CTT GCT GCT GCA ATT GAA GCA GAA	264
Thr Gln Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu	
480 485 490	
AAT CAT TGT ACA AGC CAG ACC AGA TTG CTT ATT GAC CAG ATT AGC CTG	312
Asn His Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Gln Ile Ser Leu	
495 500 505	
CAG CAA GGA AGA ATA GTT GCT CTT GAA GAA CAA ATG AAG CGT CAG GAC	360
Gln Gln Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp	
510 515 520 525	
CAG GAG TGC CGA CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG	408
Gln Glu Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys	
530 535 540	
GGC ATA AAA AAG TTG ATC GGA AAT GTA CAG ATG CCA GTG GCT GCT GTA	456
Gly Ile Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val	
545 550 555	
GTT GTT ATG GCT TGC AAT CGG GCT GAT TAC CTG GAA AAG ACT ATT AAA	504
Val Val Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys	
560 565 570	
TCC ATC TTA AAA TAC CAA ATA TCT GTT GCG TCA AAA TAT CCT CTT TTC	552
Ser Ile Leu Lys Tyr Gln Ile Ser Val Ala Ser Lys Tyr Pro Leu Phe	
575 580 585	
ATA TCC CAG GAT GGA TCA CAT CCT GAT GTC AGG AAG CTT GCT TTG AGC	600
Ile Ser Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser	
590 595 600 605	
TAT GAT CAG CTG ACG TAT ATG CAG CAC TTG GAT TTT GAA CCT GTG CAT	648
Tyr Asp Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val His	
610 615 620	
ACT GAA AGA CCA GGG GAG CTG ATT GCA TAC TAC AAA ATT GCA CGT CAT	696
Thr Glu Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His	
625 630 635	
TAC AAG TGG GCA TTG GAT CAG CTG TTT TAC AAG CAT AAT TTT AGC CGT	744
Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg	
640 645 650	
GTT ATC ATA CTA GAA GAT GAT ATG GAA ATT GCC CCT GAT TTT TTT GAC	792
Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp	
655 660 665	
TTT TTT GAG GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG	840

000030-594650

Phe 670	Phe	Glu	Ala	Gly	Ala	Thr	Leu	Leu	Asp	Arg	Asp	Lys	Ser	Ile	Met	685	
GCT Ala	ATT Ile	TCT Ser	TCT Ser	TGG Trp	AAT Asn	GAC Asp	AAT Asn	GGA Gly	CAA Gln	ATG Met	CAG Gln	TTT Phe	GTC Val	CAA Gln	GAT Asp	888	
CCT Pro	TAT Tyr	GCT Ala	CTT Leu	TAC Tyr	CGC Arg	TCA Ser	GAT Asp	TTT Phe	TTT Phe	CCC Pro	GGT Gly	CTT Leu	GGA Gly	TGG Trp	ATG Met	936	
CTT Leu	TCA Ser	AAA Lys	TCT Ser	ACT Thr	TGG Trp	GAC Asp	GAA Glu	TTA Leu	TCT Ser	CCA Pro	AAG Lys	TGG Trp	CCA Pro	AAG Lys	GCT Ala	984	
TAC Tyr	TGG Trp	GAC Asp	GAC Asp	TGG Trp	CTA Leu	AGA Arg	CTC Leu	AAA Lys	GAG Glu	AAT Asn	CAC His	AGA Arg	GGT Gly	CGA Arg	CAA Gln	1032	
TTT Phe	ATT Ile	CGC Arg	CCA Pro	GAA Glu	GTT Val	TGC Cys	AGA Arg	ACA Thr	TAT Tyr	AAT Asn	TTT Phe	GGT Gly	GAG Glu	CAT His	GGT Gly	1080	
TCT Ser	AGT Ser	TTG Leu	GGG Gly	CAG Gln	TTT Phe	TTC Phe	AAG Lys	CAG Gln	TAT Tyr	CTT Leu	GAG Glu	CCA Pro	ATT Ile	AAA Lys	CTA Leu	1128	
AAT Asn	GAT Asp	GTC Val	CAG Gln	GTT Val	GAT Asp	TGG Trp	AAG Lys	TCA Ser	ATG Met	GAC Asp	CTT Leu	AGT Ser	TAC Tyr	CTT Leu	TTG Leu	1176	
GAG Glu	GAC Asp	AAT Asn	TAC Tyr	GTG Val	AAA Lys	CAC His	TTT Phe	GGT Gly	GAC Asp	TTG Leu	GTT Val	AAA Lys	AAG Lys	GCT Ala	AAG Lys	1224	
CCC Pro	ATC Ile	CAT His	GGA Gly	GCT Ala	GAT Asp	GCT Ala	GTC Val	TTG Leu	AAA Lys	GCA Ala	TTT Phe	AAC Asn	ATA Ile	GAT Asp	GGT Gly	1272	
GAT Asp	GTG Val	CGT Arg	ATT Ile	CAG Gln	TAC Tyr	AGA Arg	GAT Asp	CAA Gln	CTA Leu	GAC Asp	TTT Phe	GAA Glu	AAT Asn	ATC Ile	GCA Ala	1320	
CGG Arg	CAA Gln	TTT Phe	GGC Gly	ATT Ile	TTT Phe	GAA Glu	GAA Glu	TGG Trp	AAG Lys	GAT Asp	GGT Gly	GTA Val	CCA Pro	CGT Arg	GCA Ala	1368	
GCA Ala	TAT Tyr	AAA Lys	GGA Gly	ATA Ile	GTA Val	GTT Val	TTC Phe	CGG Arg	TAC Tyr	CAA Gln	ACG Thr	TCC Ser	AGA Arg	CGT Arg	GTA Val	1416	
TTC Phe	CTT Leu	GTT Val	GGC Gly	CAT His	GAT Asp	TCG Ser	CTT Leu	CAA Gln	CAA Gln	CTC Leu	GGA Gly	ATT Ile	GAA Glu	GAT Asp	ACT Thr	1464	
TAA *	CAAAGATATG ATTGCAGGAG CCCGGGCCAAA ATTTTGTGACT TATTGGGTTAG															1517	
GATGCATCGA				GCTGACACTA				AACCATGATT				TTACCAGTTA				1577	
TACGGAGGAG				CTCACTGTTC				TAGTGTTGAA				GGGATATCGG				1637	

CATCAACACA ACCTATTATT TTAAGTGTTC AGAACATAAA GAGGAAATGT AGCCCTGTAA 1697
AGACTATACA TGGGACCATC ATAATCGCGG CCGCGAATTC 1737

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 447 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met	Arg	Gly	Asn	Lys	Phe	Cys	Cys	Asp	Phe	Arg	Tyr	Leu	Leu	Ile	Leu		
1				5					10					15			
Ala	Ala	Val	Ala	Phe	Ile	Tyr	Thr	Gln	Met	Arg	Leu	Phe	Ala	Thr	Gln		
			20					25					30				
Ser	Glu	Tyr	Ala	Asp	Arg	Leu	Ala	Ala	Ala	Ile	Glu	Ala	Glu	Asn	His		
			35				40					45					
Cys	Thr	Ser	Gln	Thr	Arg	Leu	Leu	Ile	Asp	Gln	Ile	Ser	Leu	Gln	Gln		
			50				55				60						
Gly	Arg	Ile	Val	Ala	Leu	Glu	Glu	Gln	Met	Lys	Arg	Gln	Asp	Gln	Glu		
					70					75					80		
Cys	Arg	Gln	Leu	Arg	Ala	Leu	Val	Gln	Asp	Leu	Glu	Ser	Lys	Gly	Ile		
				85					90					95			
Lys	Lys	Leu	Ile	Gly	Asn	Val	Gln	Met	Pro	Val	Ala	Ala	Val	Val	Val		
			100					105					110				
Met	Ala	Cys	Asn	Arg	Ala	Asp	Tyr	Leu	Glu	Lys	Thr	Ile	Lys	Ser	Ile		
		115					120					125					
Leu	Lys	Tyr	Gln	Ile	Ser	Val	Ala	Ser	Lys	Tyr	Pro	Leu	Phe	Ile	Ser		
			130				135				140						
Gln	Asp	Gly	Ser	His	Pro	Asp	Val	Arg	Lys	Leu	Ala	Leu	Ser	Tyr	Asp		
			145			150				155					160		
Gln	Leu	Thr	Tyr	Met	Gln	His	Leu	Asp	Phe	Glu	Pro	Val	His	Thr	Glu		
				165					170					175			
Arg	Pro	Gly	Glu	Leu	Ile	Ala	Tyr	Tyr	Lys	Ile	Ala	Arg	His	Tyr	Lys		
			180					185					190				
Trp	Ala	Leu	Asp	Gln	Leu	Phe	Tyr	Lys	His	Asn	Phe	Ser	Arg	Val	Ile		
		195					200					205					
Ile	Leu	Glu	Asp	Asp	Met	Glu	Ile	Ala	Pro	Asp	Phe	Phe	Asp	Phe	Phe		
		210				215					220						
Glu	Ala	Gly	Ala	Thr	Leu	Leu	Asp	Arg	Asp	Lys	Ser	Ile	Met	Ala	Ile		
					230					235				240			
Ser	Ser	Trp	Asn	Asp	Asn	Gly	Gln	Met	Gln	Phe	Val	Gln	Asp	Pro	Tyr		
			245						250					255			

Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser
260 265 270

Lys Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp
275 280 285

Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile
290 295 300

Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser
305 310 315 320

Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp
325 330 335

Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp
340 345 350

Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile
355 360 365

His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val
370 375 380

Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala Arg Gln
385 390 395 400

Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr
405 410 415

Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val Phe Leu
420 425 430

Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr *

435 440 445

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1854 base pairs
 - (B) TYPE: Nucleotide
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: No
- (iv) ANTI-SENSE: No
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Arabidopsis thaliana
 - (B) STRAIN: Columbia
 - (D) DEVELOPMENTAL STAGE: Mature plants
 - (F) TISSUE TYPE: All tissues
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: Lambda Uni-ZAP (EcoRI/XhoI) and
Lambda ACT (XhoI)
 - (B) CLONE: pBSK-Ara-GntI-full #8
- (ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1185..1193
- (D) OTHER INFORMATION:/function= "Asn Codon is a potential glycosylation site"
/product= "Consensus sequence for N-glycosylation"
/phenotype= "N glycans modulate protein characteristics"
/standard_name= "N glycosylation site"
/label= pot-CHO
/note= "absent in animal GnTI sequences"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:135..1469
- (C) IDENTIFICATION METHOD: experimental
- (D) OTHER INFORMATION:/codon_start= 135
/function= "initiates complex N glycans on secretory glycoproteins"
/EC_number= 2.4.1.101
/product= "beta-1,2-N-acetyl glucosaminyl transferase I"
/evidence= EXPERIMENTAL
/gene= "cgl"
/standard_name= "gntI"
/label= ORF
/note= "first gntI sequence from Arabidopsis (unpublished)"

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION:19..134

(ix) FEATURE:

- (A) NAME/KEY: 3'UTR
- (B) LOCATION:1470..1848

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:157..215
- (D) OTHER INFORMATION:/function= "membrane anchor (amino acids 8-27)"
/product= "hydrophobic amino-acid region in GnTI"
/standard_name= "membrane anchor of a Type II Golgi protein"
/note= "identified by comparison with animal GnTI sequences "

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1..18
- (D) OTHER INFORMATION:/function= "for preparation of a cDNA library in Lambda ACT"
/product= "XhoI-cDNA-Adaptor"
/number= 1

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1849..1854
- (D) OTHER INFORMATION:/product= "XhoI-cDNA-Adaptor"
/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGGCCA	CGAAGGCCAC	CGT	TTTTGTT	ATAACGAACG	ACACCGTTTC	AAACAAC TTC	60									
CTTATTAGCT	AGCTCCCTCC	CGG	CGGCAAA	CACCAGAAGA	TCCACCGCTT	TTGATCTGGT	120									
TGTTTGTCTG	CGAT	ATG	GCG	AGG	ATC	TCG	TGT	GAC	TTG	AGA	TTT	CTT	CTC	170		
		Met	Ala	Arg	Ile	Ser	Cys	Asp	Leu	Arg	Phe	Leu	Leu			
		1				5					10					
ATC	CCG	GCA	GCT	TTC	ATG	TTC	ATC	TAC	ATC	CAG	ATG	AGG	CTT	TTC	CAG	218
Ile	Pro	Ala	Ala	Phe	Met	Phe	Ile	Tyr	Ile	Gln	Met	Arg	Leu	Phe	Gln	
		15					20				25					
ACG	CAA	TCA	CAG	TAT	GCA	GAT	CGC	CTC	AGT	TCC	GCT	ATC	GAA	TCT	GAG	266
Thr	Gln	Ser	Gln	Tyr	Ala	Asp	Arg	Leu	Ser	Ser	Ala	Ile	Glu	Ser	Glu	
		30				35					40					
AAC	CAT	TGC	ACT	AGT	CAA	ATG	CGA	GGC	CTC	ATA	GAT	GAA	GTT	AGC	ATC	314
Asn	His	Cys	Thr	Ser	Gln	Met	Arg	Gly	Leu	Ile	Asp	Glu	Val	Ser	Ile	
		45			50					55					60	
AAA	CAG	TCG	CGG	ATT	GTT	GCC	CTC	GAA	GAT	ATG	AAG	AAC	CGC	CAG	GAC	362
Lys	Gln	Ser	Arg	Ile	Val	Ala	Leu	Glu	Asp	Met	Lys	Asn	Arg	Gln	Asp	
				65					70					75		
GAA	GAA	CTT	GTG	CAG	CTT	AAG	GAT	CTA	ATC	CAG	ACG	TTT	GAA	AAA	AAA	410
Glu	Glu	Leu	Val	Gln	Leu	Lys	Asp	Leu	Ile	Gln	Thr	Phe	Glu	Lys	Lys	
			80					85					90			
GGA	ATA	GCA	AAA	CTC	ACT	CAA	GGT	GGA	CAG	ATG	CCT	GTG	GCT	GCT	GTA	458
Gly	Ile	Ala	Lys	Leu	Thr	Gln	Gly	Gly	Gln	Met	Pro	Val	Ala	Ala	Val	
		95					100					105				
GTG	GTT	ATG	GCC	TGC	AGT	CGT	GCA	GAC	TAT	CTT	GAA	AGG	ACT	GTT	AAA	506
Val	Val	Met	Ala	Cys	Ser	Arg	Ala	Asp	Tyr	Leu	Glu	Arg	Thr	Val	Lys	
		110				115					120					
TCA	GTT	TTA	ACA	TAT	CAA	ACT	CCC	GTT	GCT	TCA	AAA	TAT	CCT	CTA	TTT	554
Ser	Val	Leu	Thr	Tyr	Gln	Thr	Pro	Val	Ala	Ser	Lys	Tyr	Pro	Leu	Phe	
					130					135					140	
ATA	TCT	CAG	GAT	GGA	TCT	GAT	CAA	GCT	GTC	AAG	AGC	AAG	TCA	TTG	AGC	602
Ile	Ser	Gln	Asp	Gly	Ser	Asp	Gln	Ala	Val	Lys	Ser	Lys	Ser	Leu	Ser	
				145					150					155		
TAT	AAT	CAA	TTA	ACA	TAT	ATG	CAG	CAC	TTG	GAT	TTT	GAA	CCA	GTG	GTC	650
Tyr	Asn	Gln	Leu	Thr	Tyr	Met	Gln	His	Leu	Asp	Phe	Glu	Pro	Val	Val	
			160					165					170			
ACT	GAA	AGG	CCT	GGT	GAA	CTG	ACT	GCG	TAC	TAC	AAG	ATT	GCA	CGT	CAC	698
Thr	Glu	Arg	Pro	Gly	Glu	Leu	Thr	Ala	Tyr	Tyr	Lys	Ile	Ala	Arg	His	
		175					180					185				
TAC	AAG	TGG	GCA	CTG	GAC	CAG	TTG	TTT	TAC	AAA	CAC	AAA	TTT	AGT	CGA	746
Tyr	Lys	Trp	Ala	Leu	Asp	Gln	Leu	Phe	Tyr	Lys	His	Lys	Phe	Ser	Arg	
		190				195					200					
GTG	ATT	ATA														

TAC TTT GAG GCT GCA GCT AGT CTC ATG GAT AGG GAT AAA ACC ATT ATG Tyr Phe Glu Ala Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met 225 230 235	842
GCT GCT TCA TCA TGG AAT GAT AAT GGA CAG AAG CAG TTT GTG CAT GAT Ala Ala Ser Ser Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp 240 245 250	890
CCC TAT GCG CTA TAC CGA TCA GAT TTT TTT CCT GGC CTT GGG TGG ATG Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met 255 260 265	938
CTC AAG AGA TCG ACT TGG GAT GAG TTA TCA CCA AAG TGG CCA AAG GCT Leu Lys Arg Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala 270 275 280	986
TAC TGG GAT GAT TGG CTG AGA CTA AAG GAA AAC CAT AAA GGC CGC CAA Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln 285 290 295 300	1034
TTC ATT GCA CCG GAA GTC TGT AGA ACA TAC AAT TTT GGT GAA CAT GGG Phe Ile Ala Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly 305 310 315	1082
TCT AGT TTG GGA CAG TTT TTC AGT CAG TAT CTG GAA CCT ATA AAG CTA Ser Ser Leu Gly Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu 320 325 330	1130
AAC GAT GTG ACG GTT GAC TGG AAA GCA AAG GAC CTG GGA TAC CTG ACA Asn Asp Val Thr Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr 335 340 345	1178
GAG GGA AAC TAT ACC AAG TAC TTT TCT GGC TTA GTG AGA CAA GCA CGA Glu Gly Asn Tyr Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg 350 355 360	1226
CCA ATT CAA GGT TCT GAC CTT GTC TTA AAG GCT CAA AAC ATA AAG GAT Pro Ile Gln Gly Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp 365 370 375 380	1274
GAT GAT CGT ATC CGG TAT AAA GAC CAA GTA GAG TTT GAA CGC ATT GCA Asp Asp Arg Ile Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala 385 390 395	1322
GGG GAA TTT GGT ATA TTT GAA GAA TGG AAG GAT GGT GTG CCA CGA ACA Gly Glu Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr 400 405 410	1370
GCA TAT AAA GGA GTA GTG GTG TTT CGA ATC CAG ACA ACA AGA CGT GTA Ala Tyr Lys Gly Val Val Val Phe Arg Ile Gln Thr Thr Arg Arg Val 415 420 425	1418
TTC CTG GTT GGG CCA GAT TCT GTA ATG CAG CTT GGA ATT CGA AAT TCC Phe Leu Val Gly Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser 430 435 440	1466
TGA TGCAAAACAT ATGAAAGGAA AAGAAGATTT TGGACCGCAT GCAGCCTCCT * 445	1519
TCTAGCAGCT GTTAGGTTGT ATTGTTATTT ATGGATGAGT TTGTAGAGCG GTGGGGTTAA	1579
CTTTAACAGC AAGGAAGCTC TGGTGACCAG GCTGATTGGC TTAGAAGTTA TGGGAACCCC	1639

000000-999999

TTGAAAGGGT CAGGGTTAAA TATATTTTCAG TTGTTTTTATT AGTGATTATC TTGTGGGTAA 1699
 CTTATACGAA TGCAAATCAT TCTATGCAGT TTTTCTTCGT CCCACTTGTT TTGGCTTCTC 1759
 TATTGCTAGT GTACATATCT CTTCAAACAT GACTAAATA ATGCGTGTTG CTTCAAAGAA 1819
 GTAACTTTTA TTAAAAA AAAA AAAAAC TCGAG 1854

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 445 amino acids
- (B) TYPE: Amino acid
- (D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met	Ala	Arg	Ile	Ser	Cys	Asp	Leu	Arg	Phe	Leu	Leu	Ile	Pro	Ala	Ala	1	5	10	15
Phe	Met	Phe	Ile	Tyr	Ile	Gln	Met	Arg	Leu	Phe	Gln	Thr	Gln	Ser	Gln	20	25	30	
Tyr	Ala	Asp	Arg	Leu	Ser	Ser	Ala	Ile	Glu	Ser	Glu	Asn	His	Cys	Thr	35	40	45	
Ser	Gln	Met	Arg	Gly	Leu	Ile	Asp	Glu	Val	Ser	Ile	Lys	Gln	Ser	Arg	50	55	60	
Ile	Val	Ala	Leu	Glu	Asp	Met	Lys	Asn	Arg	Gln	Asp	Glu	Glu	Leu	Val	65	70	75	80
Gln	Leu	Lys	Asp	Leu	Ile	Gln	Thr	Phe	Glu	Lys	Lys	Gly	Ile	Ala	Lys	85	90	95	
Leu	Thr	Gln	Gly	Gly	Gln	Met	Pro	Val	Ala	Ala	Val	Val	Val	Met	Ala	100	105	110	
Cys	Ser	Arg	Ala	Asp	Tyr	Leu	Glu	Arg	Thr	Val	Lys	Ser	Val	Leu	Thr	115	120	125	
Tyr	Gln	Thr	Pro	Val	Ala	Ser	Lys	Tyr	Pro	Leu	Phe	Ile	Ser	Gln	Asp	130	135	140	
Gly	Ser	Asp	Gln	Ala	Val	Lys	Ser	Lys	Ser	Leu	Ser	Tyr	Asn	Gln	Leu	145	150	155	160
Thr	Tyr	Met	Gln	His	Leu	Asp	Phe	Glu	Pro	Val	Val	Thr	Glu	Arg	Pro	165	170	175	
Gly	Glu	Leu	Thr	Ala	Tyr	Tyr	Lys	Ile	Ala	Arg	His	Tyr	Lys	Trp	Ala	180	185	190	
Leu	Asp	Gln	Leu	Phe	Tyr	Lys	His	Lys	Phe	Ser	Arg	Val	Ile	Ile	Leu	195	200	205	
Glu	Asp	Asp	Met	Glu	Ile	Ala	Pro	Asp	Phe	Phe	Asp	Tyr	Phe	Glu	Ala	210	215	220	
Ala	Ala	Ser	Leu	Met	Asp	Arg	Asp	Lys	Thr	Ile	Met	Ala	Ala	Ser	Ser				

000000-999999

225		230		235		240
Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp Pro Tyr Ala Leu	245		250		255	
Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Lys Arg Ser	260		265		270	
Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp Asp	275		280		285	
Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln Phe Ile Ala Pro	290		295		300	
Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu Gly	305		310		315	320
Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val Thr	325		330		335	
Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr Glu Gly Asn Tyr	340		345		350	
Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg Pro Ile Gln Gly	355		360		365	
Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp Asp Asp Arg Ile	370		375		380	
Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala Gly Glu Phe Gly	385		390		395	400
Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr Ala Tyr Lys Gly	405		410		415	
Val Val Val Phe Arg Ile Gln Thr Thr Arg Arg Val Phe Leu Val Gly	420		425		430	
Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser *	435		440		445	

C L A I M S

1. Method for the production of glycoproteins displaying minimal, uniform and defined sugar residues, comprising cultivating a transgenic plant, parts of transgenic plants or transformed plant cells, and isolating the desired glycoprotein from the material cultivated, characterized in that the transgenic plant, parts of transgenic plants or transformed plant cells, respectively, is/are transformed with an antisense construct or a sense construct, comprising an antisense DNA or a sense DNA with respect to the DNA sequence for a gene or a cDNA for plant N-acetyl glucosaminyl transferase I or a part thereof, for elimination or reduction of the activity of said N-acetyl glucosaminyl transferase, wherein the antisense or sense construct optionally contains additional regulatory sequences for the transcription of the respective antisense or sense DNA.
2. Method according to claim 1, characterized in that for transformation an antisense or sense construct with respect to one of the cDNAs encoding N-acetyl glucosaminyl transferase I from *Solanum tuberosum*, *Nicotiana tabacum* or *Arabidopsis thaliana* is used.
3. Method according to claim 2, characterized in that for transformation an antisense or sense construct with respect to one of the DNA sequences given in SEQ ID NO: 1, 3 or 5 is used.
4. Method according to any of the claims 1 to 3, characterized in that the transgenic plant used is

additionally transformed with the gene encoding the desired glycoprotein.

- 5 5. DNA, characterized in that it encodes N-acetyl glucosaminyl transferase I from *Solanum tuberosum*.
- 10 6. DNA according to claim 5, characterized in that it comprises the nucleotide sequence given in SEQ ID NO: 1 or a part thereof.
- 15 7. DNA, characterized in that it encodes N-acetyl glucosaminyl transferase I from *Nicotiana tabacum*.
- 20 8. DNA according to claim 7, characterized in that it comprises the nucleotide sequence given in SEQ ID NO: 3 or a part thereof.
- 25 9. DNA encoding N-acetyl glucosaminyl transferase I from *Arabidopsis thaliana*, characterized in that said DNA encodes the amino-acid sequence given in SEQ ID NO: 6 or the nucleotide sequence given in SEQ ID NO: 5 or a part thereof.
- 30 10. DNA, characterized in that it comprises the nucleotide sequence complementary to the DNA according to claim 6, 8 or 9.
- 35 11. DNA, characterized in that it may be obtained by substitution, deletion and/or insertion of one or more nucleotides and/or truncation at the 5' and/or 3' end of one of the DNAs according to any of the claims 5 to 10, with the proviso, that said DNA hybridizes at least in a partial region with the starting DNA or its complementary sequence or parts thereof under stringent conditions.
- 40

12. DNA, characterized in that it represents a gene or is part of a gene, which encodes the enzyme N-acetyl glucosaminyl transferase I, and which in its entirety or in a partial region thereof hybridizes under stringent conditions

- to one of the DNA sequences or fragments according to any of the claims 5 to 11 and/or

- to a DNA sequence, which has been derived from the amino acid sequences given in SEQ ID NO: 1, 3 and/or 5, considering the degeneration of the genetic code.

13. DNA construct, characterized in that it comprises one or more of the DNAs according to any of the claims 5 to 14.

14. DNA construct according to claim 13, characterized in that it comprises an antisense or sense DNA with respect to the DNA sequence according to any of the claims 5 to 12 and optionally regulatory sequences for the transcription of the antisense or sense DNA, respectively.

15. Vector, plasmid, cosmid, virus or phage genome, characterized in that it contains at least a DNA and/or construct according to any of the claims 5 to 14.

16. N-acetyl glucosaminyl transferase I from *Solanum tuberosum*.

17. N-acetyl glucosaminyl transferase I from *Nicotiana tabacum*.

18. N-acetyl glucosaminyl transferase I from *Arabidopsis thaliana*, characterized in that the enzyme comprises the amino acid sequence set forth in SEQ ID NO: 6.

19. N-acetyl glucosaminyl transferase I, characterized in that the enzyme comprises the amino acid sequence set forth in SEQ ID NO: 2.

20. N-acetyl glucosaminyl transferase I, characterized in that the enzyme comprises amino acids 74 to 446 of the amino acid sequence set forth in SEQ ID NO: 2.

21. N-acetyl glucosaminyl transferase I, characterized in that the enzyme comprises the amino acid sequence set forth in SEQ ID NO: 4.

22. N-acetyl glucosaminyl transferase I, available due to hybridization of its gene or one or more of the portions of its gene to one or more of the DNAs and/or DNA fragments according to any of the claims 5 to 12.

23. Enzymes or proteins derived from the enzymes according to any of the claims 16 to 22 by substitution, deletion, insertion and/or modification of individual amino acids and/or smaller groups of amino acids and/or by N- and/or C-terminal truncation and/or extension.

24. Protein or peptide, comprising one or more portions of the amino acid sequence(s) of one or more of the enzymes defined in any of the claims 16 to 23.

25. Protein or peptide, encoded by one of the DNAs according to any of the claims 5 to 12.

26. Antigen, characterized in that it comprises:
- the amino acid sequence given in SEQ ID NO: 2, SEQ ID NO: 4 or SEQ ID NO: 6, or
- amino acids 74 to 446 of the amino acid sequence given in Fig. 2, or
- an amino acid sequence derived from the amino acid sequences given in SEQ ID NO: 2, 4 or 6 by substitution,

- deletion, insertion and/or modification of individual amino acids and/or smaller groups of amino acids, or
- one or more parts of said sequences,
with the proviso, that upon immunization of a host with the antigen, said antigen may raise an immunological reaction, including the production of antibodies directed against the antigen.
27. Monoclonal or polyclonal antibody, characterized in that it specifically recognizes and binds one or more of the enzymes or antigens according to any of the claims 16 to 26.
28. Microorganism,
characterized in that it is transformed by at least one of the nucleotide sequences selected from the DNAs, constructs, vectors, plasmids, cosmids, virus or phage genomes according to one or more of the claims 5 to 15.
29. Transgenic plant, transgenic seed, transgenic reproduction material, parts of transgenic plants or transformed plant cell, obtainable by integration of one or more DNA sequence(s) or construct(s) according to any of the claims 5 to 13 under the control of a promoter effective in plants, into the genome of a plant, or via infection by means of a virus containing one or more DNA sequence(s) or construct(s) according to any of the claims 5 to 13, for an extrachromosomal propagation and expression of the DNA sequence(s) or construct(s) in the plant tissue infected.
30. Transgenic plant, transgenic seed, transgenic reproduction material, parts of transgenic plants or transformed plant cell with missing or reduced N-acetyl glucosaminyl transferase I activity, obtainable by integration of one or more antisense or sense construct(s) according to claim 14 under the control of a promoter

effective in plants, into the genome of a plant, or by
viral infection by means of a virus containing one or
more antisense or sense construct(s) according to claim
14, for an extrachromosomal propagation and
transcription of the antisense construct(s) in the plant
tissue infected.

ABSTRACT OF THE DISCLOSURE

5 This invention relates to plant *GntI* sequences, in particular
to plant nucleic acid sequences encoding the enzyme N-acetyl
glucosaminyl transferase I (GnTI), DNA sequences derived
therefrom, including *GntI* antisense and sense constructs, and
the translation products thereof, antibodies directed against
10 said translation products, as well as the use of the sequence
information for the production of transformed microorganisms
and transgenic plants, including those having reduced or
missing N-acetyl glucosaminyl transferase I activity. Such
15 plants displaying reduced or lacking N-acetyl glucosaminyl
transferase I activity are of great importance for the pro-
duction of glycoproteins of specific constitution with
respect to their sugar residues.

20
25

30

35

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: von Schaewen, Antje Dr. rer. nat.
- (B) STREET: Natruperstrasse 169a
- (C) CITY: Osnabrueck
- (E) COUNTRY: Germany
- (F) POSTAL CODE (ZIP): D-49076
- (G) TELEPHONE: +49-541-684029

(ii) TITLE OF INVENTION: Plant gntI sequences and the use thereof for the production of plants having reduced or lacking N-acetyl glucosaminyl transferase I (GnTI) activity

(iii) NUMBER OF SEQUENCES: 6

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1669 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Solanum tuberosum
- (B) STRAIN: Desiree
- (D) DEVELOPMENTAL STAGE: Sink organ
- (F) TISSUE TYPE: Mesophyll
- (G) CELL TYPE: Leaf cells

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: Lambda ZAP II (Eco RI)
- (B) CLONE: gntI-A1(K)

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 659..667
- (D) OTHER INFORMATION: /function= "Asn codon in this context is a potential glycosylation site"
/product= "N-glycosylation consensus sequence"
/phenotype= "N-glycans modulate protein properties"
/standard_name= "N-glycosylation site"
/label= pot-CHO
/note= "GnTI-coding sequences from animals do not contain this feature"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 53..1393
- (C) IDENTIFICATION METHOD: experimental

(D) OTHER INFORMATION:/codon_start= 53
 /function= "initiates complex N-glycans on
 secretory glycoproteins"
 /EC_number= 2.4.1.101
 /product=
 "beta-1,2-N-acetylglucosaminyltransferase I"
 /evidence= EXPERIMENTAL
 /gene= "cgl"
 /standard_name= "gntI"
 /label= ORF
 /note= "first gntI sequence from potato (unpublished)"

(ix) FEATURE:
 (A) NAME/KEY: 5'UTR
 (B) LOCATION:15..52

(ix) FEATURE:
 (A) NAME/KEY: 3'UTR
 (B) LOCATION:1394..1655

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION:80..139
 (D) OTHER INFORMATION:/function= "membrane anchor (amino
 acids 10-29)"
 /product= "hydrophobic amino acid stretch in GntI"
 /standard_name= "membrane anchor of a type II
 Golgi protein"
 /note= "identified by comparison with GntI sequences
 from animals"

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:1..14
 (D) OTHER INFORMATION:/function= "used for cloning the
 cDNA library in Lambda ZAPII"
 /product= "EcoRI/NotI-cDNA adapter"
 /number= 1

(ix) FEATURE:
 (A) NAME/KEY: misc_feature
 (B) LOCATION:1656..1669
 (D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
 /number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAATTCGCGG CCGCCTGAGA AACCTCGAA TTCAATTTTCG CATTTGGCAG AG ATG	55
Met	
1	
AGA GGG AAC AAG TTT TGC TTT GAT TTA CGG TAC CTT CTC GTC GTG GCT	103
Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val Ala	
5 10 15	
GCT CTC GCC TTC ATC TAC ATA CAG ATG CGG CTT TTC GCG ACA CAG TCA	151
Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln Ser	
20 25 30	
GAA TAT GTA GAC CGC CTT GCT GCT GCA ATT GAA GCA GAA AAT CAT TGT	199
Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His Cys	
35 40 45	
ACA AGT CAG ACC AGA TTG CTT ATT GAC AAG ATT AGC CAG CAG CAA GGA	247

Thr	Ser	Gln	Thr	Arg	Leu	Leu	Ile	Asp	Lys	Ile	Ser	Gln	Gln	Gln	Gly		
50					55					60					65		
AGA	GTA	GTA	GCT	CTT	GAA	GAA	CAA	ATG	AAG	CAT	CAG	GAC	CAG	GAG	TGC	295	
Arg	Val	Val	Ala	Leu	Glu	Glu	Gln	Met	Lys	His	Gln	Asp	Gln	Glu	Cys		
				70					75					80			
CGG	CAA	TTA	AGG	GCT	CTT	GTT	CAG	GAT	CTT	GAA	AGT	AAG	GGC	ATA	AAA	343	
Arg	Gln	Leu	Arg	Ala	Leu	Val	Gln	Asp	Leu	Glu	Ser	Lys	Gly	Ile	Lys		
			85					90					95				
AAG	TTA	ATC	GGA	GAT	GTG	CAG	ATG	CCA	GTG	GCA	GCT	GTA	GTT	GTT	ATG	391	
Lys	Leu	Ile	Gly	Asp	Val	Gln	Met	Pro	Val	Ala	Ala	Val	Val	Val	Met		
		100					105					110					
GCT	TGC	AGT	CGT	ACT	GAC	TAC	CTG	GAG	AGG	ACT	ATT	AAA	TCC	ATC	TTA	439	
Ala	Cys	Ser	Arg	Thr	Asp	Tyr	Leu	Glu	Arg	Thr	Ile	Lys	Ser	Ile	Leu		
			115			120					125						
AAA	TAC	CAA	ACA	TCT	GTT	GCA	TCA	AAA	TAT	CCT	CTT	TTC	ATA	TCC	CAG	487	
Lys	Tyr	Gln	Thr	Ser	Val	Ala	Ser	Lys	Tyr	Pro	Leu	Phe	Ile	Ser	Gln		
					135					140					145		
GAT	GGA	TCA	AAT	CCT	GAT	GTA	AGA	AAG	CTT	GCT	TTG	AGC	TAT	GGT	CAG	535	
Asp	Gly	Ser	Asn	Pro	Asp	Val	Arg	Lys	Leu	Ala	Leu	Ser	Tyr	Gly	Gln		
				150					155					160			
CTG	ACG	TAT	ATG	CAG	CAC	TTG	GAT	TAT	GAA	CCT	GTG	CAT	ACT	GAA	AGA	583	
Leu	Thr	Tyr	Met	Gln	His	Leu	Asp	Tyr	Glu	Pro	Val	His	Thr	Glu	Arg		
			165					170					175				
CCA	GGG	GAA	CTG	GTT	GCA	TAC	TAC	AAG	ATT	GCA	CGT	CAT	TAC	AAG	TGG	631	
Pro	Gly	Glu	Leu	Val	Ala	Tyr	Tyr	Lys	Ile	Ala	Arg	His	Tyr	Lys	Trp		
			180					185				190					
GCA	TTG	GAT	CAG	CTG	TTT	CAC	AAG	CAT	AAT	TTT	AGC	CGT	GTT	ATC	ATA	679	
Ala	Leu	Asp	Gln	Leu	Phe	His	Lys	His	Asn	Phe	Ser	Arg	Val	Ile	Ile		
			195			200				205							
CTA	GAA	GAT	GAT	ATG	GAA	ATT	GCT	GCT	GAT	TTT	TTT	GAC	TAT	TTT	GAG	727	
Leu	Glu	Asp	Asp	Met	Glu	Ile	Ala	Ala	Asp	Phe	Phe	Asp	Tyr	Phe	Glu		
					215					220					225		
GCT	GGA	GCT	ACT	CTT	CTT	GAC	AGA	GAC	AAG	TCG	ATT	ATG	GCT	ATT	TCT	775	
Ala	Gly	Ala	Thr	Leu	Leu	Asp	Arg	Asp	Lys	Ser	Ile	Met	Ala	Ile	Ser		
				230					235					240			
TCT	TGG	AAT	GAC	AAT	GGA	CAA	AGG	CAG	TTC	GTC	CAA	GAT	CCT	GAT	GCT	823	
Ser	Trp	Asn	Asp	Asn	Gly	Gln	Arg	Gln	Phe	Val	Gln	Asp	Pro	Asp	Ala		
			245					250					255				

000090-9946500

CTT TAC CGC TCA GAC TTT TTT CCT GGT CTT GGA TGG ATG CTT TCA AAA	871
Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser Lys	
260 265 270	
TCA ACT TGG TCC GAA CTA TCT CCA AAG TGG CCA AAG GCT TAC TGG GAT	919
Ser Thr Trp Ser Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp	
275 280 285	
GAC TGG CTA AGG CTG AAA GAA AAT CAC AGA GGT CGA CAA TTT ATT CGC	967
Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile Arg	
290 295 300 305	
CCA GAA GTT TGC AGA ACG TAC AAT TTT GGT GAG CAT GGT TCT AGT TTG	1015
Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu	
310 315 320	
GGG CAG TTT TTT AAG CAG TAT CTT GAG CCA ATT AAG CTA AAT GAT GTC	1063
Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val	
325 330 335	
CAG GTT GAT TGG AAG TCA ATG GAC CTA AGT TAC CTT TTG GAG GAC AAC	1111
Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp Asn	
340 345 350	
TAT GTG AAA CAC TTT GGC GAC TTG GTT AAA AAG GCT AAG CCC ATC CAC	1159
Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile His	
355 360 365	
GGA GCT GAT GCT GTT TTG AAA GCA TTT AAC ATA GAT GGT GAT GTG CGT	1207
Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val Arg	
370 375 380 385	
ATT CAG TAC AGA GAC CAA CTA GAC TTT GAA GAT ATC GCT CGA CAG TTT	1255
Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln Phe	
390 395 400	
GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGG GCA GCA TAT AAA	1303
Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr Lys	
405 410 415	
GGG ATA GTA GTT TTC CGG TTT CAA ACA TCT AGA CGT GTG TTC CTT GTT	1351
Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu Val	
420 425 430	
TCC CCT GAT TCT CTT CGA CAA CTT GGA GTT GAA GAT ACT TAG	1393
Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr *	
435 440 445	
CGAAGATATG ATTGGAGCCT GAGCAACAAT TTAGACTTAT TTGGTAGGAT ACATTTGAAA	1453
GAGCTGACAC GAAAAGTATG ACTACCAGTA GCTACATGCA ACATTTTAAT GTTAATGGAA	1513
GGAACCCACT GCTTATTGTT GGAATGGATG AATCATCACC ACATCCTATT ATTCAAGTTT	1573
ACAAACATAA AGAGGAAATG TTGCCCTATA AAAACAAATT TTTTGTCTTCT AAGAAGGAAC	1633
GTTACGATTA TGAGCAACTT TGGCGGCCGC GAATTC	1669

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87													

(A) LENGTH: 447 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met 1	Arg	Gly	Asn	Lys 5	Phe	Cys	Phe	Asp	Leu 10	Arg	Tyr	Leu	Leu	Val 15	Val
Ala	Ala	Leu	Ala 20	Phe	Ile	Tyr	Ile	Gln 25	Met	Arg	Leu	Phe	Ala 30	Thr	Gln
Ser	Glu	Tyr 35	Val	Asp	Arg	Leu	Ala 40	Ala	Ala	Ile	Glu	Ala 45	Glu	Asn	His
Cys	Thr 50	Ser	Gln	Thr	Arg	Leu 55	Leu	Ile	Asp	Lys	Ile 60	Ser	Gln	Gln	Gln
Gly 65	Arg	Val	Val	Ala	Leu 70	Glu	Glu	Gln	Met	Lys 75	His	Gln	Asp	Gln	Glu 80
Cys	Arg	Gln	Leu	Arg 85	Ala	Leu	Val	Gln	Asp 90	Leu	Glu	Ser	Lys	Gly 95	Ile
Lys	Lys	Leu	Ile 100	Gly	Asp	Val	Gln	Met 105	Pro	Val	Ala	Ala	Val 110	Val	Val
Met	Ala	Cys 115	Ser	Arg	Thr	Asp	Tyr 120	Leu	Glu	Arg	Thr	Ile 125	Lys	Ser	Ile
Leu	Lys 130	Tyr	Gln	Thr	Ser	Val 135	Ala	Ser	Lys	Tyr	Pro 140	Leu	Phe	Ile	Ser
Gln 145	Asp	Gly	Ser	Asn	Pro 150	Asp	Val	Arg	Lys	Leu 155	Ala	Leu	Ser	Tyr	Gly 160
Gln	Leu	Thr	Tyr	Met 165	Gln	His	Leu	Asp	Tyr 170	Glu	Pro	Val	His	Thr 175	Glu
Arg	Pro	Gly	Glu 180	Leu	Val	Ala	Tyr	Tyr 185	Lys	Ile	Ala	Arg	His 190	Tyr	Lys
Trp	Ala	Leu 195	Asp	Gln	Leu	Phe	His 200	Lys	His	Asn	Phe	Ser 205	Arg	Val	Ile
Ile	Leu 210	Glu	Asp	Asp	Met	Glu 215	Ile	Ala	Ala	Asp	Phe 220	Phe	Asp	Tyr	Phe
Glu 225	Ala	Gly	Ala	Thr	Leu 230	Leu	Asp	Arg	Asp	Lys 235	Ser	Ile	Met	Ala	Ile 240
Ser	Ser	Trp	Asn	Asp 245	Asn	Gly	Gln	Arg	Gln	Phe	Val	Gln	Asp	Pro 255	Asp
Ala	Leu	Tyr	Arg 260	Ser	Asp	Phe	Phe	Pro 265	Gly	Leu	Gly	Trp	Met 270	Leu	Ser

Lys	Ser	Thr	Trp	Ser	Glu	Leu	Ser	Pro	Lys	Trp	Pro	Lys	Ala	Tyr	Trp
		275					280					285			
Asp	Asp	Trp	Leu	Arg	Leu	Lys	Glu	Asn	His	Arg	Gly	Arg	Gln	Phe	Ile
	290					295					300				
Arg	Pro	Glu	Val	Cys	Arg	Thr	Tyr	Asn	Phe	Gly	Glu	His	Gly	Ser	Ser
305				310						315					320
Leu	Gly	Gln	Phe	Phe	Lys	Gln	Tyr	Leu	Glu	Pro	Ile	Lys	Leu	Asn	Asp
				325					330					335	
Val	Gln	Val	Asp	Trp	Lys	Ser	Met	Asp	Leu	Ser	Tyr	Leu	Leu	Glu	Asp
			340					345					350		
Asn	Tyr	Val	Lys	His	Phe	Gly	Asp	Leu	Val	Lys	Lys	Ala	Lys	Pro	Ile
		355					360					365			
His	Gly	Ala	Asp	Ala	Val	Leu	Lys	Ala	Phe	Asn	Ile	Asp	Gly	Asp	Val
	370					375					380				
Arg	Ile	Gln	Tyr	Arg	Asp	Gln	Leu	Asp	Phe	Glu	Asp	Ile	Ala	Arg	Gln
385					390					395					400
Phe	Gly	Ile	Phe	Glu	Glu	Trp	Lys	Asp	Gly	Val	Pro	Arg	Ala	Ala	Tyr
				405					410					415	
Lys	Gly	Ile	Val	Val	Phe	Arg	Phe	Gln	Thr	Ser	Arg	Arg	Val	Phe	Leu
			420					425					430		
Val	Ser	Pro	Asp	Ser	Leu	Arg	Gln	Leu	Gly	Val	Glu	Asp	Thr	*	
		435					440					445			

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1737 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(B) STRAIN: Samsun NN

(F) TISSUE TYPE: Mesophyll

(vii) IMMEDIATE SOURCE:

(B) CLONE: qntI-A9 (T)

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:733..741
- (D) OTHER INFORMATION:/function= "Asn codon in this context is a potential glycosylation site"
/product= "N-glycosylation consensus sequence"
/phenotype= "N-glycans modulate protein properties"
/standard_name= "N-glycosylation site"
/label= pot-CHO
/note= "GnTI sequences from animals do not contain this feature"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:127..1467
- (C) IDENTIFICATION METHOD: experimental
- (D) OTHER INFORMATION:/codon_start= 127
/function= "initiates complex N-glycans on secretory glycoproteins"
/EC_number= 2.4.1.101
/product= "beta-1,2-N-acetylglucosaminyltransferase I"
/evidence= EXPERIMENTAL
/gene= "cgl"
/standard_name= "gntI"
/label= ORF
/note= "first gntI sequence from tobacco (unpublished)"

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION:15..126

(ix) FEATURE:

- (A) NAME/KEY: 3'UTR
- (B) LOCATION:1468..1723

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:154..213
- (D) OTHER INFORMATION:/function= "membrane anchor (amino acids 10-29)"
/product= "hydrophobic amino acid stretch in GnTI"
/standard_name= "membrane anchor of a type II golgi protein"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1..14
- (D) OTHER INFORMATION:/function= "use for cloning the cDNA library in Lambda ZAPII"
/product= "EcoRI/NotI-cDNA adapter"
/number= 1

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION:1724..1737
- (D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GAATTCGCGG CCGCCATTGA CTTGATCCTA ACTGAACAGG CAAAGTAAAT CCAGCGATGA	60
AACACTCATA ACTGAACACT GAGAGACTAT TCGCTTTCTC CTAAAGCCTT CAATCGAATT	120
CGCACG ATG AGA GGG AAC AAG TTT TGC TGT GAT TTC CGG TAC CTC CTC Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu	168
450 455 460	
ATC TTG GCT GCT GTC GCC TTC ATC TAC ACA CAG ATG CGG CTT TTT GCG Ile Leu Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala	216
465 470 475	
ACA CAG TCA GAA TAT GCA GAT CGC CTT GCT GCT GCA ATT GAA GCA GAA Thr Gln Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu	264
480 485 490	
AAT CAT TGT ACA AGC CAG ACC AGA TTG CTT ATT GAC CAG ATT AGC CTG Asn His Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Gln Ile Ser Leu	312
495 500 505	
CAG CAA GGA AGA ATA GTT GCT CTT GAA GAA CAA ATG AAG CGT CAG GAC Gln Gln Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp	360
510 515 520 525	
CAG GAG TGC CGA CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG Gln Glu Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys	408
530 535 540	
GGC ATA AAA AAG TTG ATC GGA AAT GTA CAG ATG CCA GTG GCT GCT GTA Gly Ile Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val	456
545 550 555	
GTT GTT ATG GCT TGC AAT CGG GCT GAT TAC CTG GAA AAG ACT ATT AAA Val Val Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys	504
560 565 570	
TCC ATC TTA AAA TAC CAA ATA TCT GTT GCG TCA AAA TAT CCT CTT TTC Ser Ile Leu Lys Tyr Gln Ile Ser Val Ala Ser Lys Tyr Pro Leu Phe	552
575 580 585	
ATA TCC CAG GAT GGA TCA CAT CCT GAT GTC AGG AAG CTT GCT TTG AGC Ile Ser Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser	600
590 595 600 605	
TAT GAT CAG CTG ACG TAT ATG CAG CAC TTG GAT TTT GAA CCT GTG CAT Tyr Asp Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val His	648
610 615 620	
ACT GAA AGA CCA GGG GAG CTG ATT GCA TAC TAC AAA ATT GCA CGT CAT Thr Glu Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His	696
625 630 635	
TAC AAG TGG GCA TTG GAT CAG CTG TTT TAC AAG CAT AAT TTT AGC CGT Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg	744
640 645 650	
GTT ATC ATA CTA GAA GAT GAT ATG GAA ATT GCC CCT GAT TTT TTT GAC Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp	792
655 660 665	
TTT TTT GAG GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG Phe Phe Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met	840
670 675 680 685	

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 447 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu Ile Leu
1 5 10 15
Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala Thr Gln
20 25 30
Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His
35 40 45
Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Gln Ile Ser Leu Gln Gln
50 55 60
Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp Gln Glu
65 70 75 80
Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile
85 90 95
Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val Val Val
100 105 110
Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys Ser Ile
115 120 125
Leu Lys Tyr Gln Ile Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser
130 135 140
Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Asp
145 150 155 160
Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val His Thr Glu
165 170 175
Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys
180 185 190
Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg Val Ile
195 200 205
Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp Phe Phe
210 215 220
Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile
225 230 235 240
Ser Ser Trp Asn Asp Asn Gly Gln Met Gln Phe Val Gln Asp Pro Tyr
245 250 255
Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser
260 265 270
Lys Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp
275 280 285
Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile

290	295	300
Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser 305 310 315 320		
Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp 325 330 335		
Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp 340 345 350		
Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile 355 360 365		
His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val 370 375 380		
Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala Arg Gln 385 390 395 400		
Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr 405 410 415		
Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val Phe Leu 420 425 430		
Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr *		
435 440 445		

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1854 base pairs
 - (B) TYPE: Nucleotide
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: No
- (iv) ANTI-SENSE: No
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Arabidopsis thaliana
 - (B) STRAIN: Columbia
 - (D) DEVELOPMENTAL STAGE: Mature plants
 - (F) TISSUE TYPE: All tissues
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: Lambda Uni-ZAP (EcoRI/XhoI) and Lambda ACT (XhoI)
 - (B) CLONE: pBSK-Ara-GntI-full #8
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 1185..1193
 - (D) OTHER INFORMATION: /function= "Asn Codon is a potential glycosylation site"
 - /product= "Consensus sequence for N-glycosylation"
 - /phenotype= "N glycans modulate protein characteristics"
 - /standard_name= "N glycosylation site"

/label= pot-CHO
/note= "absent in animal GnTI sequences"

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:135..1469
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION:/codon_start= 135
/function= "initiates complex N glycans on
secretory glycoproteins"
/EC_number= 2.4.1.101
/product=
"beta-1,2-N-acetyl glucosaminyl transferase I"
/evidence= EXPERIMENTAL
/gene= "cgl"
/standard_name= "gntI"
/label= ORF
/note= "first gntI sequence from Arabidopsis
(unpublished)"

(ix) FEATURE:

(A) NAME/KEY: 5'UTR
(B) LOCATION:19..134

(ix) FEATURE:

(A) NAME/KEY: 3'UTR
(B) LOCATION:1470..1848

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:157..215
(D) OTHER INFORMATION:/function= "membrane anchor
(amino acids 8-27)"
/product= "hydrophobic amino-acid region in
GnTI"
/standard_name= "membrane anchor of a Type II
Golgi protein"
/note= "identified by comparison with animal GnTI
sequences "

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1..18
(D) OTHER INFORMATION:/function= "for preparation
of a cDNA library in Lambda ACT"
/product= "XhoI-cDNA-Adaptor"
/number= 1

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION:1849..1854
(D) OTHER INFORMATION:/product= "XhoI-cDNA-Adaptor"
/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGGCCA CGAAGGCCAC CGTTTTTGTGTT ATAACGAACG ACACCGTTTC AAACAACCTC	60
CTTATTAGCT AGCTCCCTCC CGGCGGCAAA CACCAGAAGA TCCACCGCTT TTGATCTGGT	120
TGTTTGTCGT CGAT ATG GCG AGG ATC TCG TGT GAC TTG AGA TTT CTT CTC	170
Met Ala Arg Ile Ser Cys Asp Leu Arg Phe Leu Leu	
1 5 10	

Leu	Lys	Arg	Ser	Thr	Trp	Asp	Glu	Leu	Ser	Pro	Lys	Trp	Pro	Lys	Ala		
270						275					280						
TAC	TGG	GAT	GAT	TGG	CTG	AGA	CTA	AAG	GAA	AAC	CAT	AAA	GGC	CGC	CAA	1034	
Tyr	Trp	Asp	Asp	Trp	Leu	Arg	Leu	Lys	Glu	Asn	His	Lys	Gly	Arg	Gln		
285					290				295						300		
TTC	ATT	GCA	CCG	GAA	GTC	TGT	AGA	ACA	TAC	AAT	TTT	GGT	GAA	CAT	GGG	1082	
Phe	Ile	Ala	Pro	Glu	Val	Cys	Arg	Thr	Tyr	Asn	Phe	Gly	Glu	His	Gly		
				305					310					315			
TCT	AGT	TTG	GGA	CAG	TTT	TTC	AGT	CAG	TAT	CTG	GAA	CCT	ATA	AAG	CTA	1130	
Ser	Ser	Leu	Gly	Gln	Phe	Phe	Ser	Gln	Tyr	Leu	Glu	Pro	Ile	Lys	Leu		
			320					325					330				
AAC	GAT	GTG	ACG	GTT	GAC	TGG	AAA	GCA	AAG	GAC	CTG	GGA	TAC	CTG	ACA	1178	
Asn	Asp	Val	Thr	Val	Asp	Trp	Lys	Ala	Lys	Asp	Leu	Gly	Tyr	Leu	Thr		
		335					340					345					
GAG	GGA	AAC	TAT	ACC	AAG	TAC	TTT	TCT	GGC	TTA	GTG	AGA	CAA	GCA	CGA	1226	
Glu	Gly	Asn	Tyr	Thr	Lys	Tyr	Phe	Ser	Gly	Leu	Val	Arg	Gln	Ala	Arg		
	350					355					360						
CCA	ATT	CAA	GGT	TCT	GAC	CTT	GTC	TTA	AAG	GCT	CAA	AAC	ATA	AAG	GAT	1274	
Pro	Ile	Gln	Gly	Ser	Asp	Leu	Val	Leu	Lys	Ala	Gln	Asn	Ile	Lys	Asp		
365					370				375						380		
GAT	GAT	CGT	ATC	CGG	TAT	AAA	GAC	CAA	GTA	GAG	TTT	GAA	CGC	ATT	GCA	1322	
Asp	Asp	Arg	Ile	Arg	Tyr	Lys	Asp	Gln	Val	Glu	Phe	Glu	Arg	Ile	Ala		
				385					390					395			
GGG	GAA	TTT	GGT	ATA	TTT	GAA	GAA	TGG	AAG	GAT	GGT	GTG	CCA	CGA	ACA	1370	
Gly	Glu	Phe	Gly	Ile	Phe	Glu	Glu	Trp	Lys	Asp	Gly	Val	Pro	Arg	Thr		
			400					405					410				
GCA	TAT	AAA	GGA	GTA	GTG	GTG	TTT	CGA	ATC	CAG	ACA	ACA	AGA	CGT	GTA	1418	
Ala	Tyr	Lys	Gly	Val	Val	Val	Phe	Arg	Ile	Gln	Thr	Thr	Arg	Arg	Val		
		415					420					425					
TTC	CTG	GTT	GGG	CCA	GAT	TCT	GTA	ATG	CAG	CTT	GGA	ATT	CGA	AAT	TCC	1466	
Phe	Leu	Val	Gly	Pro	Asp	Ser	Val	Met	Gln	Leu	Gly	Ile	Arg	Asn	Ser		
	430					435					440						
TGA	TGCAAAACAT	ATGAAAGGAA	AAGAAGATTT	TGGACCGCAT	GCAGCCTCCT											1519	
*																	
445																	
TCTAGCAGCT	GTTAGGTTGT	ATTGTTATTT	ATGGATGAGT	TTGTAGAGCG	GTGGGGTTAA											1579	
CTTTAACAGC	AAGGAAGCTC	TGGTGACCAG	GCTGATTGGC	TTAGAAGTTA	TGGGAACCCC											1639	
TTGAAAGGGT	CAGGGTTAAA	TATATTTT	CAGTTGTTTT	TATTATC	TTGTGGGTAA											1699	
CTTATACGAA	TGCAAATCAT	TCTATGCAGT	TTTTCTTCGT	CCCACTTGTT	TTGGCTTCTC											1759	
TATTGCTAGT	GTACATATCT	CTTCAAACAT	GTACTAAATA	ATGCGTGTTG	CTTCAAAGAA											1819	
GTAACCTTTTA	TTAAAAAAA	AAAAAAAAC	TCGAG													1854	

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 445 amino acids
 - (B) TYPE: Amino acid

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Ala Arg Ile Ser Cys Asp Leu Arg Phe Leu Leu Ile Pro Ala Ala
1 5 10 15
Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln Thr Gln Ser Gln
20 25 30
Tyr Ala Asp Arg Leu Ser Ser Ala Ile Glu Ser Glu Asn His Cys Thr
35 40 45
Ser Gln Met Arg Gly Leu Ile Asp Glu Val Ser Ile Lys Gln Ser Arg
50 55 60
Ile Val Ala Leu Glu Asp Met Lys Asn Arg Gln Asp Glu Glu Leu Val
65 70 75 80
Gln Leu Lys Asp Leu Ile Gln Thr Phe Glu Lys Lys Gly Ile Ala Lys
85 90 95
Leu Thr Gln Gly Gly Gln Met Pro Val Ala Ala Val Val Val Met Ala
100 105 110
Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys Ser Val Leu Thr
115 120 125
Tyr Gln Thr Pro Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln Asp
130 135 140
Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser Tyr Asn Gln Leu
145 150 155 160
Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val Val Thr Glu Arg Pro
165 170 175
Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp Ala
180 185 190
Leu Asp Gln Leu Phe Tyr Lys His Lys Phe Ser Arg Val Ile Ile Leu
195 200 205
Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp Tyr Phe Glu Ala
210 215 220
Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met Ala Ala Ser Ser
225 230 235 240
Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp Pro Tyr Ala Leu
245 250 255
Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Lys Arg Ser
260 265 270
Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp Asp
275 280 285
Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln Phe Ile Ala Pro
290 295 300
Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu Gly
305 310 315 320

000090-9946500

Figure 1

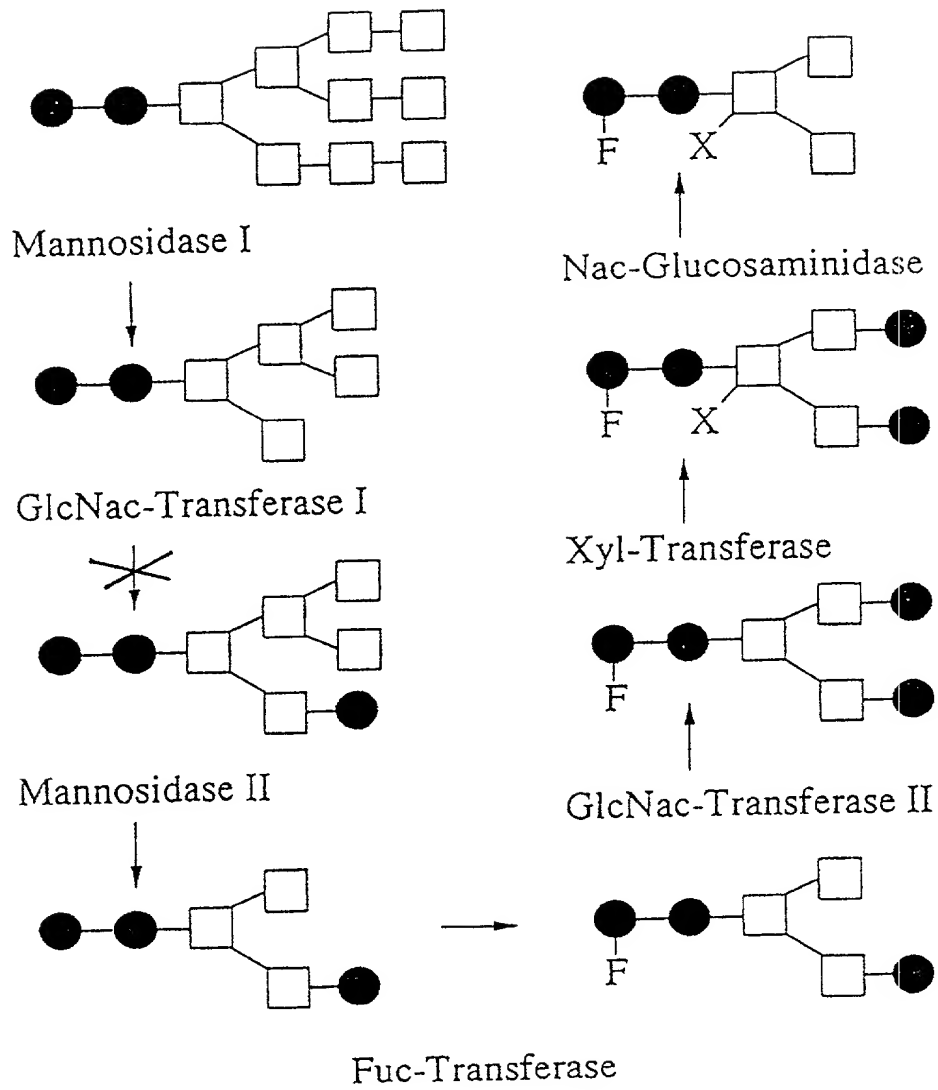


Figure 2

Al *GntI* cDNA

GAATTCGCGG CCGCCTGAGA AACCCCTCGAA TTCAATTTTCG CATTTGGCAG AG ATG	55
Met	
1	
AGA GGG AAC AAG TTT TGC TTT GAT TTA CGG TAC CTT CTC GTC GTG GCT	103
Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val Ala	
5 10 15	
GCT CTC GCC TTC ATC TAC ATA CAG ATG CGG CTT TTC GCG ACA CAG TCA	151
Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln Ser	
20 25 30	
GAA TAT GTA GAC CGC CTT GCT GCT GCA ATT GAA GCA GAA AAT CAT TGT	199
Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His Cys	
35 40 45	
ACA AGT CAG ACC AGA TTG CTT ATT GAC AAG ATT AGC CAG CAG CAA GGA	247
Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gln Gly	
50 55 60 65	
AGA GTA GTA GCT CTT GAA GAA CAA ATG AAG CAT CAG GAC CAG GAG TGC	295
Arg Val Val Ala Leu Glu Glu Gln Met Lys His Gln Asp Gln Glu Cys	
70 75 80	
CGG CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG GGC ATA AAA	343
Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile Lys	
85 90 95	
AAG TTA ATC GGA GAT GTG CAG ATG CCA GTG GCA GCT GTA GTT GTT ATG	391
Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val Met	
100 105 110	
GCT TGC AGT CGT ACT GAC TAC CTG GAG AGG ACT ATT AAA TCC ATC TTA	439
Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile Leu	
115 120 125	
AAA TAC CAA ACA TCT GTT GCA TCA AAA TAT CCT CTT TTC ATA TCC CAG	487
Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln	
130 135 140 145	
GAT GGA TCA AAT CCT GAT GTA AGA AAG CTT GCT TTG AGC TAT GGT CAG	535
Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly Gln	
150 155 160	
CTG ACG TAT ATG CAG CAC TTG GAT TAT GAA CCT GTG CAT ACT GAA AGA	583
Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu Arg	
165 170 175	
CCA GGG GAA CTG GTT GCA TAC TAC AAG ATT GCA CGT CAT TAC AAG TGG	631
Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp	
180 185 190	
GCA TTG GAT CAG CTG TTT CAC AAG CAT AAT TTT AGC CGT GTT ATC ATA	679
Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile Ile	
195 200 205	
CTA GAA GAT GAT ATG GAA ATT GCT GCT GAT TTT TTT GAC TAT TTT GAG	727
Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe Glu	
210 215 220 225	

Figure 2 (continued)

GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG <u>GCT ATT TCT</u>	775
Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile Ser	
230 235 240	
<u>TCT TGG AAT GAC AAT GGA CAA AGG CAG TTC GTC CAA GAT CCT GAT GCT</u>	823
Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp Ala	
245 250 255	
CTT TAC CGC TCA <u>GAC TTT TTT CCT GGT CTT GGA TGG</u> ATG CTT TCA AAA	871
Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser Lys	
260 265 270	
TCA ACT TGG TCC GAA CTA TCT CCA AAG TGG CCA AAG GCT TAC TGG GAT	919
Ser Thr Trp Ser Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp	
275 280 285	
GAC TGG CTA AGG CTG AAA GAA AAT CAC AGA GGT CGA CAA TTT ATT CGC	967
Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile Arg	
290 295 300 305	
CCA GAA GTT TGC AGA ACG TAC AAT TTT GGT GAG CAT GGT TCT AGT TTG	1015
Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu	
310 315 320	
GGG CAG TTT TTT AAG CAG TAT CTT GAG CCA ATT AAG CTA AAT GAT GTC	1063
Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val	
325 330 335	
CAG GTT GAT TGG AAG TCA ATG GAC CTA AGT TAC CTT TTG GAG GAC AAC	1111
Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp Asn	
340 345 350	
TAT GTG AAA CAC TTT GGC GAC TTG GTT AAA AAG GCT AAG CCC ATC CAC	1159
Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile His	
355 360 365	
GGA GCT GAT GCT GTT TTG AAA GCA TTT AAC ATA GAT GGT GAT GTG CGT	1207
Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val Arg	
370 375 380 385	
ATT CAG TAC AGA GAC CAA CTA GAC TTT GAA GAT ATC GCT CGA CAG TTT	1255
Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln Phe	
390 395 400	
GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGG GCA GCA TAT AAA	1303
Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr Lys	
405 410 415	
GGG ATA GTA GTT TTC CGG TTT CAA ACA TCT AGA CGT GTG TTC CTT GTT	1351
Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu Val	
420 425 430	
TCC CCT GAT TCT CTT CGA CAA CTT GGA GTT GAA GAT ACT TAG	1393
Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr End	
435 440 445	
CGAAGATATG ATTGGAGCCT GAGCAACAAT TTAGACTTAT TTGGTAGGAT ACATTTGAAA	1453
GAGCTGACAC GAAAAGTATG ACTACCAGTA GCTACATGCA ACATTTTAAAT GTTAATGGAA	1513
GGAACCCACT GCTTATTGTT GGAATGGATG AATCATCACC ACATCCTATT ATTCAAGTTT	1573
ACAAACATAA AGAGGAAATG TTGCCCTATA AAAACAAATT TTTTGTTTCT AAGAAGGAAC	1633
GTTACGATTA TGAGCAACTT TGGCGGCCGC GAATTC	1669

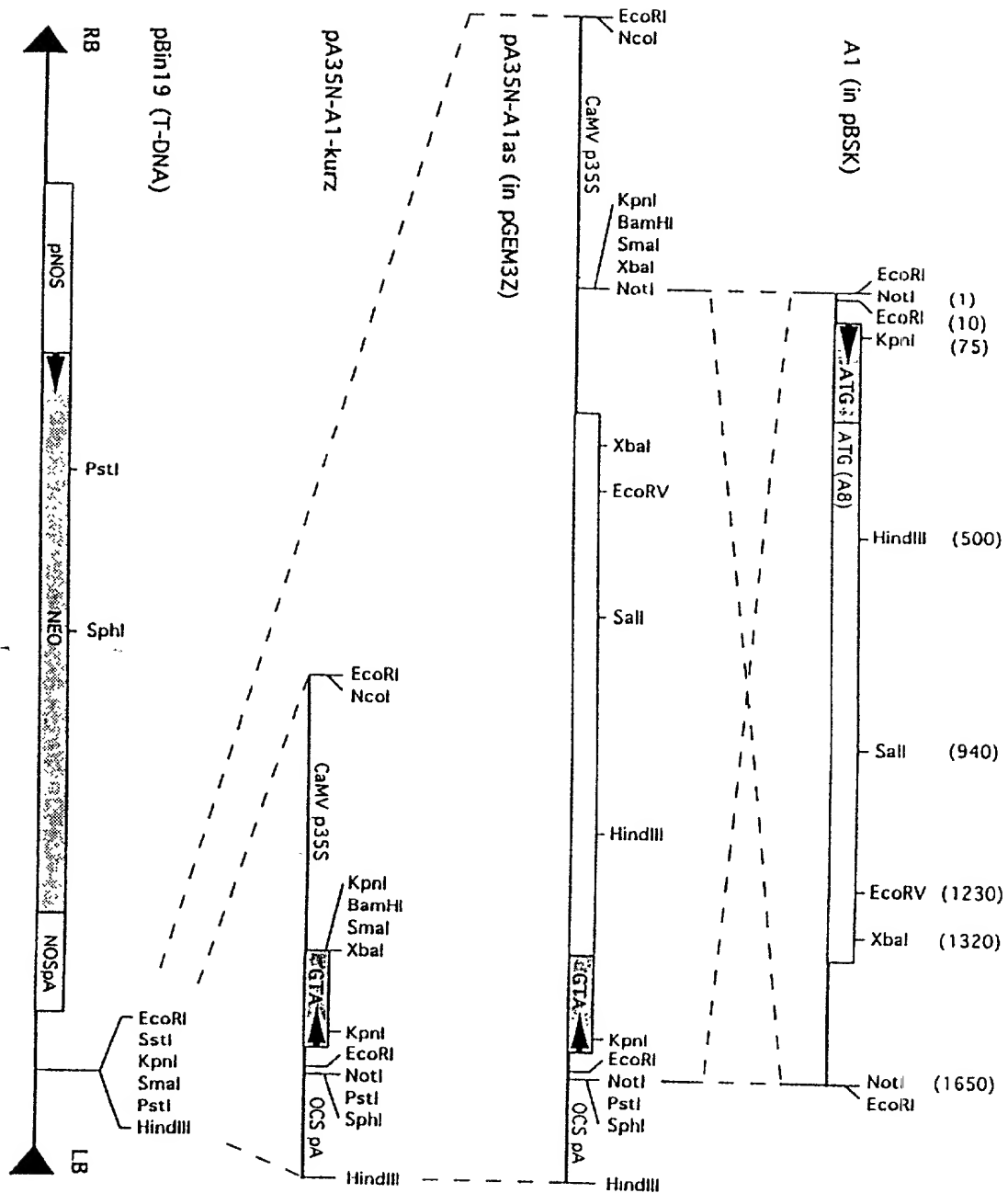
Figure 3A

Hu	Ra	Mo	Ce	St
35 (59)	36 (57)	35 (59)	33 (57)	
	92 (95)	91 (94)	38 (57)	Hu
		90 (93)	38 (57)	Ra
			38 (58)	Mo

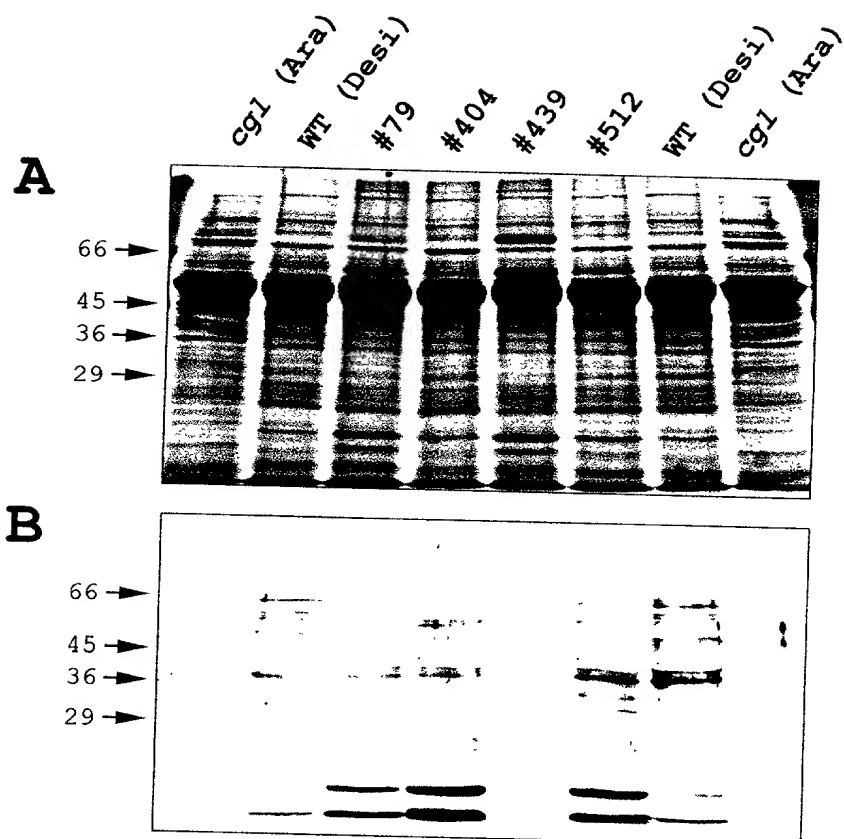
Figure 3B

A_Stb-A1	1	MRGNKECEDLRYLLAAAFIYIOMRLFATQSEYVDRLAAAEAEHCT
B_Ntb-A9	1	MRGNKECCDFRYLLIAAFIYIOMRLFATQSEYADRLAAAEAEHCT
C_Atb-Full	1	..MAFISCDLRLLIPAAFMFIYIOMRLFOTOSVADRLESAESENHCT
A_Stb-A1	51	SQTRLLIDKISQOGRVALEEOMKQDQECROLRALVODLESKGIKKLI
B_Ntb-A9	51	SQTRLLIDKISQOGRVALEEOMKQDQECROLRALVODLESKGIKKLI
C_Atb-Full	49	SQMRGLIDKISQOGRVALEEOMKQDQECROLRALVODLESKGIKKLI
A_Stb-A1	101	GVVOMPVAADVVMACSRDYLERTIKSILKYOTSVASKYPLFISODGSGNP
B_Ntb-A9	101	GVVOMPVAADVVMACSRDYLERTIKSILKYOTSVASKYPLFISODGSGHP
C_Atb-Full	99	QGGOMPVAADVVMACSRDYLERTIKSILKYOTSVASKYPLFISODGSGQ
A_Stb-A1	151	DVRKLALSVCOLTYMOHLDPEVHTERPGELTAYYKIARHYKWALDOLFEH
B_Ntb-A9	151	DVRKLALSVCOLTYMOHLDPEVHTERPGELTAYYKIARHYKWALDOLFFY
C_Atb-Full	149	AVNSKLSVCOLTYMOHLDPEVHTERPGELTAYYKIARHYKWALDOLFFY
A_Stb-A1	201	KHNFSRVIILEDDMEIAADFFDYFEAGATLLDRDKSIMAISSWNDNGOMQO
B_Ntb-A9	201	KHNFSRVIILEDDMEIAADFFDYFEAGATLLDRDKSIMAISSWNDNGOMQO
C_Atb-Full	199	KHKFSRVIILEDDMEIAADFFDYFEAGATLLDRDKSIMAISSWNDNGOMQO
A_Stb-A1	251	FVQDPDALYRSDFEPGLGWMLSKSTWSELSPKWPKEYWDDWLRLKENHRG
B_Ntb-A9	251	FVQDPYALYRSDFEPGLGWMLSKSTWDELSPKWPKEYWDDWLRLKENHRG
C_Atb-Full	249	FVQDPYALYRSDFEPGLGWMLSKSTWDELSPKWPKEYWDDWLRLKENHRG
A_Stb-A1	301	ROFIRPEVCRITYNFEHGSSLGQFFKOYLEPIKLNVDVQVWKSMDLSYLL
B_Ntb-A9	301	ROFIRPEVCRITYNFEHGSSLGQFFKOYLEPIKLNVDVQVWKSMDLSYLL
C_Atb-Full	299	ROFIAPVCRITYNFEHGSSLGQFFKOYLEPIKLNVDVQVWKSMDLSYLL
A_Stb-A1	351	EDNYVKHFGDLVKKAKPIHGADAVLKAQFNIDGDVRIQYRDOLDFFDIARQ
B_Ntb-A9	351	EDNYVKHFGDLVKKAKPIHGADAVLKAQFNIDGDVRIQYRDOLDFFDIARQ
C_Atb-Full	349	EGNYTKYFSGLVQAPPIQGGDLVLKAQFNIDGDVRIQYRDOLDFFDIARQ
A_Stb-A1	401	FGIFEWKDGVPRAAAYKGIVVFRQTSRRVFLVSPDSLRLGLGIEDT
B_Ntb-A9	401	FGIFEWKDGVPRAAAYKGIVVFRQTSRRVFLVSPDSLRLGLGIEDT
C_Atb-Full	399	FGIFEWKDGVPRAAAYKGIVVFRQTSRRVFLVSPDSLRLGLGIEDT

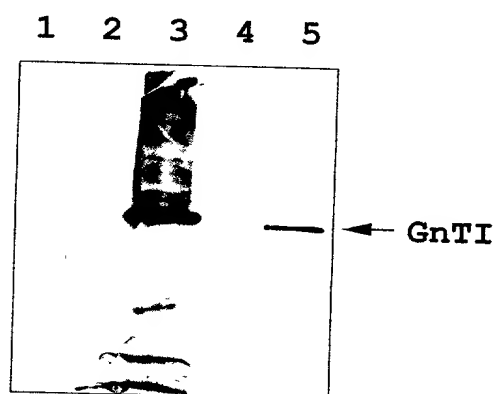
Figure 4



Figur 5



Figur 6



COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY
(Includes Reference to Provisional and PCT International Applications)

Attorney's Docket No.

032266-003

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name;

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

Plant GntI sequences and the use thereof for the production of plants having reduced or lacking N-acetyl
glucosaminyl transferase I (GnTI) activity

the specification of which (check only one item below):

☐ is attached hereto.

☒ was filed as United States application

Number Unassigned

on June 9, 2000

and was amended

on June 9, 2000 (if applicable).

☐ was filed as PCT international application

Number _____

on _____

and was amended

on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 (a)-(e) of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. §119:

COUNTRY (if PCT, indicate "PCT")	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED UNDER 35 U.S.C. §119
Germany	197 54 622.6	09 12 1997	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
			<input type="checkbox"/> Yes <input type="checkbox"/> No
			<input type="checkbox"/> Yes <input type="checkbox"/> No
			<input type="checkbox"/> Yes <input type="checkbox"/> No
			<input type="checkbox"/> Yes <input type="checkbox"/> No

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below.

(Application Number)

(Filing Date)

(Application Number)

(Filing Date)

COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (CONT'D)
(Includes Reference to Provisional and PCT International Applications)

Attorney's Docket No.

032266-003

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose to the Office all information known to me to be material to the patentability as defined in Title 37, Code of Federal Regulations §1.56, which became available between the filing date of the prior application(s) and the national or PCT international filing date of this application:

PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS DESIGNATING THE U.S. FOR BENEFIT UNDER 35 U.S.C. §120:

U.S. APPLICATIONS		STATUS (check one)		
U.S. APPLICATION NUMBER	U.S. FILING DATE	PATENTED	PENDING	ABANDONED
PCT APPLICATIONS DESIGNATING THE U.S.				
PCT APPLICATION NO.	PCT FILING DATE	U.S. APPLICATION NUMBERS ASSIGNED (if any)		
EP98/08001	09 12 98			

I hereby appoint the following attorneys and agent(s) to prosecute said application and to transact all business in the Patent and Trademark Office connected therewith and to file, prosecute and to transact all business in connection with international applications directed to said invention:

William L. Mathis	17,337	R. Danny Huntington	27,903	Gerald F. Swiss	30,113
Robert S. Swecker	19,885	Eric H. Weisblatt	30,505	Michael J. Ure	33,089
Platon N. Mandros	22,124	James W. Peterson	26,057	Charles F. Wieland III	33,096
Benton S. Duffett, Jr.	22,030	Teresa Stanek Rea	30,427	Bruce T. Wieder	33,815
Norman H. Stepno	22,716	Robert E. Krebs	25,885	Todd R. Walters	34,040
Ronald L. Grudziecki	24,970	William C. Rowland	30,888	Ronni S. Jillions	31,979
Frederick G. Michaud, Jr.	26,003	T. Gene Dillahunty	25,423	Harold R. Brown III	36,341
Alan E. Kopecki	25,813	Patrick C. Keane	32,858	Allen R. Baum	36,086
Regis E. Slutter	26,999	Bruce J. Boggs, Jr.	32,344	Steven M. du Bois	35,023
Samuel C. Miller, III	27,360	William H. Benz	25,952	Brian P. O'Shaughnessy	32,747
Robert G. Mukai	28,531	Peter K. Skiff	31,917	Kenneth B. Leffler	36,075
George A. Hovanec, Jr.	28,223	Richard J. McGrath	29,195	Fred W. Hathaway	32,236
James A. LaBarre	28,632	Matthew L. Schneider	32,814		
E. Joseph Gess	28,510	Michael G. Savage	32,596		



21839

and: Ping Hwung and Leslie A. Mooi

Address all correspondence to:



21839

William H. Benz
BURNS, DOANE, SWECKER & MATHIS, L.L.P.
P.O. Box 1404
Alexandria, Virginia 22313-1404

Address all telephone calls to: William H. Benz at (650) 622-2300.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (CONT'D)
(Includes Reference to Provisional and PCT International Applications)

Attorney's Docket No.

032266-003

FULL NAME OF SOLE OR FIRST INVENTOR	SIGNATURE	DATE
ANTJE VON SCHAEWEN		
RESIDENCE	CITIZENSHIP	
NATRUPER STRAßE 169A, 49076 OSNABRÜCK GERMANY		GERMAN
POST OFFICE ADDRESS		
FULL NAME OF SECOND JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		
CITIZENSHIP		
POST OFFICE ADDRESS		
FULL NAME OF THIRD JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		
CITIZENSHIP		
POST OFFICE ADDRESS		
FULL NAME OF FOURTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		
CITIZENSHIP		
POST OFFICE ADDRESS		
FULL NAME OF FIFTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		
CITIZENSHIP		
POST OFFICE ADDRESS		
FULL NAME OF SIXTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		
CITIZENSHIP		
POST OFFICE ADDRESS		
FULL NAME OF SEVENTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		
CITIZENSHIP		
POST OFFICE ADDRESS		
FULL NAME OF EIGHTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		
CITIZENSHIP		
POST OFFICE ADDRESS		